Three Essays in Macroeconometrics

Inaugural-Dissertation
zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften
durch die
Rechts- und Staatswissenschaftliche Fakultät
der Rheinischen Friedrich-Wilhelms-Universität
Bonn

vorgelegt von

Sinem Hacıoğlu Hoke

aus Ankara (Türkei)

Bonn 2016

Dekan:	Prof. Dr. Rainer Hüttemann
Erstreferent:	Prof. Dr. Matei Demetrescu
Zweitreferent:	Prof. Dr. Christian Pigorsch

Tag der mündlichen Prüfung: 27.09.2016

Acknowledgements

This dissertation would not have been completed without the help and the support of many people. First and foremost, I would like to express my gratitude to my supervisor Matei Demetrescu who has stood by me since my first day. His sincere comments and endless help enabled me to come this far. He was always more than a supervisor for me. I appreciate what he taught me as a professor and a mentor.

I would like to thank my second advisor Christian Pigorsch for believing in me and agreeing to supervise my second paper even though I was abroad. I am very grateful to find the opportunity of having him in my committee. Many thanks to Benjamin Born for being the third member of my thesis committee.

I am thankful to be given the opportunity to be a part of Bonn Graduate School of Economics. I would also like to thank the German Research Foundation (DFG) for financial support in this respect. I was very privileged to go to University of Wisconsin-Madison for a year during my studies. I could not have done it without the aid of and the financial support through the International Office. I greatly benefitted from the fruitful comments of Jack Porter.

Many thanks go to my dear friend and my coauthor Kerem Tuzcuoglu for supporting me, helping me a great deal and even covering my lack of proceeding on certain concepts. I know that he will be a great academic one day and more people will benefit from his sight. I will be there to give him a big applause.

I would like to thank Jeremy Chiu to support me during my internship at the Bank of England. His support made it possible me to come where I am now. I will be always most thankful to your help on the third chapter of this thesis. Moreover, I am very lucky to have the opportunity to work with Tomasz Wieladek during the internship.

Life would have been really difficult without my friends in Bonn. I established life time friendships here. I am indebted to Buket and Demet for sharing both my problems and my joy with me in the last 6 years. They always believed in me, looked out for me and opened their priceless hearts to me. You will be present in every single stage of my life.

My fellow grad friends, Andreas Kleiner, Anke Becker, Benjamin Enke, Désirée Rückert, Lien Pham-Dao, Moritz Drexl, Ralph Lütticke created a great atmosphere despite the hard processes we had been all going through. Living in a different country and trying to get used to a new life along with professional obligations were simplified by your existence in my life. I will definitely miss watching soccer, having beer and chatting with you. I am very lucky to have you all. I am looking forward to dancing at more weddings! And I am very lucky to be the aunt of Ben and Anke's daughter, the cutest Leni ever!

Most importantly, special thanks to my family who supported me along this long path. Mom, you have been always there for me even through Skype. You always made me feel that I have a safe place to be. Gizem, you are the best sister ever and I still have no idea how you are always able to cheer me up! Dad, I know you are there somewhere watching me.

Lastly, I would like to thank my Justin for showing up in my life unexpectedly, very surprisingly but exultantly. I will always appreciate the circumstances which lead me your way. I could not have survived the last three years without the thought of our amazing future. I know you will be always there for me.

Sinem Hacioglu Hoke
London
2016

Introduction

Motivated by the recent availability of extensive macroeconomic data sets, this thesis consists of three independent chapters that examine the ways to approach to this issue from various angles.

CHAPTER 1. The first chapter, which is a joint work with Matei Demetrescu, discusses the particularities of forecasting with factor-augmented predictive regressions under general loss functions. In line with the literature, principal component analysis is employed to extract factors from the set of predictors. We also extract information on the volatility of the series to be predicted, since volatility is forecast-relevant under non-quadratic loss functions. Moreover, the predictive regression is estimated by minimizing the in-sample average loss, to ensure asymptotic unbiasedness of forecasts under the relevant loss. Finally, to select the most promising predictors for the series to be forecast, we employ an information criterion tailored to the relevant loss. Using the Stock and Watson data set, we assess the proposed adjustments in a pseudo out-of-sample forecasting exercise. Expectedly, the use of estimation under the relevant loss is found to be effective. In other words, the forecasting exercises we employ suggest that evaluating forecasts under the chosen asymmetric loss function lead to smaller forecast losses. Using an additional volatility proxy as predictor and conducting model selection tailored to the relevant loss function further enhance forecasts. Both the theoretical and the empirical results emphasize the importance of using the relevant loss functions while performing forecasting exercises with extracted factors.

CHAPTER 2. The second chapter is a joint paper with Kerem Tuzcuoglu and is linked to the first chapter in the factor analysis sense. Researches have not found a way to assign economic meaning to factors although factor analysis has been widely used as a dimension reduction method. In this paper, we propose a Threshold Factor Augmented Vector Autoregression model to address this issue. The novelty is the interpretation of factors by observing how frequently factor loadings fall below estimated thresholds and become irrelevant. The results indicate that we are able to relate most of the factors to specific categories of the data without any prior specification on the data set. Furthermore, the interpretable factors, e.g., real activity factor, unemployment factor and such, are each given shocks along with policy shock to observe the responses of the other factors and individual series. The resulting impulse responses are of expected sign and magnitude in general. Overall, our results yield an intensive layout on which factor is associated with different aspects of the economy. We are able to associate the extracted factors with certain macroeconomic activities, such as real economy, unemployment, inflation. We present impulse response analysis to show how a contractionary monetary shock affects the factors and variables. Our results are consistent with what the economic theory suggests.

CHAPTER 3. The third chapter, written in collaborative work with Jeremy Chiu, uses the large information sets in vector autoregression sense. Motivated by the desire to probe macroeconomic tail events and to capture nonlinear economic dynamics, we estimate two types of regime switching models with Bayesian estimation methods: Threshold VAR and Markov switching VAR. We also use linear Bayesian VAR model as a benchmark. For each of the non-linear models, we estimate regimes which carry the interpretation of recessionary/normal and financially stressful/stable periods. Using the recursiveness assumption and conditional on shocks of one-standard-deviation, we show that (i) *financial shocks* hitting during times of recessions create disproportionately more severe contractions in output; (ii) *output growth shocks* hitting in financially stressful times result in disproportionately further financial stress. We also demonstrate the power of a *feedback loop* between real and financial sectors when extremely large shocks hit the economy in normal/financially stable periods. Afterwards, we perform out-of-sample forecasting exercises, and find that the Threshold VAR model has the potential to predict tail events in *conditional* forecasting compared to the Markov switching VAR and Bayesian VAR. Our findings provide strong evidence of nonlinearities and shock amplification mechanisms in the United Kingdom data, and hence useful information to investigate macroeconomic tail events.

Contents

Introduction . . . iii

1 *Macroeconomic Forecasting with Large Data Sets under Asymmetric Loss* **1**

1.1 Motivation . . . 1

1.2 The basic forecasting problem . . . 4

1.3 Extracting additional relevant information . . . 7

1.4 Forecasting under asymmetric loss . . . 11

1.4.1 Setup . . . 11

1.4.2 Extracted factors . . . 12

1.4.3 Results . . . 15

1.5 Concluding remarks . . . 18

Appendix 1 **20**

1.A An information criterion . . . 20

1.B Proof of Proposition 1 . . . 20

1.C Model selection using the LASSO . . . 23

1.D Data description . . . 29

2 *Interpreting Latent Dynamic Factors by Threshold FAVAR Model* **33**

2.1 Introduction 33

2.2 The Model 36

2.2.1 Model Specification 36

2.2.2 Bayesian Estimation 38

2.2.3 Identification Restrictions for Factors 39

2.3 The Data 40

2.4 Results 41

2.4.1 Number of Factors in the Data 41

2.4.2 Interpreting the Factors 42

2.5 Impulse Response Functions 45

2.6 Concluding Remarks 48

Appendix 2 **51**

2.A Priors and Posteriors 51

2.A.1 Priors 51

2.A.2 MCMC Estimation Steps 51

2.B Different Restrictions on Λ_t 56

2.C Impulse response functions 57

2.D Data Description For SW Updated Data 60

3 *Macroeconomic Tail Events with Non-linear BVARs* **64**

3.1 Introduction 64

3.2 Model Specifications 68

3.2.1 Threshold VAR 69

3.2.2 Markov Switching Vector Autoregression Models 70

3.2.3 Benchmark Bayesian VAR 70

3.3 Data 70

3.4 Full sample estimation results 71

3.5 Impulse Response Analysis of Structural Shocks 73

3.5.1 Generalised impulse response functions and shock identification 74

3.5.2 Shock propagation and macro tail events 75

3.5.3 Tail shocks and the feedback loop 78

3.6 Forecasting analysis 79

3.7 Concluding remarks 80

Appendix 3 **90**

3.A Normal Inverse Wishart priors 90

3.B The Gibbs Sampling algorithm for TVAR 91

3.C The Gibbs Sampling algorithm for MSVAR 91

3.C.1 An extension: MSVAR with two independent Markov chains 93

3.D Generalized Impulse Response Functions 95

3.D.1 GIRFs for TVAR 95

3.D.2 GIRFs for MSVAR 96

3.E Variables 96

Bibliography **99**

List of Figures

2.1 The correlation between the variables and the first factor 44

2.2 The responses of the factors and FFR to a 1 unit shock on FFR. 46

2.3 The responses of the factors and FFR to a 1 unit shock on unemployment factor. 47

2.4 The responses of the variables to a 1 unit shock on *FFR*. 47

2.5 The responses of the factors and FFR to a 1 unit shock on *inflation and finance* factor. 57

2.6 The responses of the factors and FFR to a 1 unit shock on *third* factor. 57

2.7 The responses of the factors and FFR to a 1 unit shock on *real activity* factor. . 58

2.8 The responses of the factors and FFR to a 1 unit shock on *expectations* factor. . 58

2.9 The responses of the factors and FFR to a 1 unit shock on *money* factor. 59

2.10 The responses of the factors and FFR to a 1 unit shock on *interest rate* factor. . 59

3.1 Full sample regimes for TVAR-Y 72

3.2 Full sample regimes for TVAR-S 73

3.3 Full sample regime probabilities for MSVAR 74

3.4 Impulse Responses to a 1 SD adverse shock to corporate bond spreads in BVAR model . . . 82

3.5 Generalised Impulse Responses to a 1 SD adverse shock to corporate bond spreads in TVAR-Y model . . . 82

3.6 Generalised Impulse Responses to a 1 SD adverse shock to corporate bond spreads in TVAR-S model . . . 83

3.7 Generalised Impulse Responses to a 1 SD adverse shock to corporate bond spreads in MSVAR model . . . 83

3.8 Impulse Responses to a 1 SD adverse shock to real GDP growth in BVAR model 84

3.9 Generalised Impulse Responses to a 1 SD adverse shock to real GDP growth in TVAR-Y model . . . 84

3.10 Generalised Impulse Responses to a 1 SD adverse shock to real GDP growth in TVAR-S model . . . 85

3.11 Generalised Impulse Responses to a 1 SD adverse shock to real GDP growth in MSVAR model . . . 85

3.12 Impulse Responses to a 1 SD adverse shock to short term interest rate in BVAR model . . . 86

3.13 Generalised Impulse Responses to a 1 SD adverse shock to short term interest rate in TVAR-Y model . . . 86

3.14 Generalised Impulse Responses to a **3 SD adverse shock to real GDP growth** in TVAR-Y model (to be compared with Figure 3.9) . . . 87

3.15 Generalised Impulse Responses to a **3 SD adverse shock to corporate bond spreads** in TVAR-S model (to be compared with Figure 3.6) . . . 87

3.16 Unconditional predictive densities for the three models . . . 88

3.17 Conditional predictive densities based on the path of Corporate Bond Spreads . . 89

3.1 Full sample regimes for MSVAR with 2 independent Markov chains . . . 94

3.2 Variables . . . 98

List of Tables

1.1 Factors selected for predicting for all predictor series by AIC$_{\mathcal{L}}$; full data span . . 14

1.2 Losses Evaluated for OLS-Asymmetric Loss and Asymmetric Loss 16

1.3 Tests of equal predictive accuracy of OLS and AL based forecasts 17

1.4 Selection of Factors (LASSO), full data span 25

1.5 Average Losses with LASSO-based model selection - Parsimonious post-fit LASSO 27

1.6 Average Losses with LASSO-based model selection - Flexible post-fit LASSO . . 28

2.1 Subgroups in the Data Set 41

2.2 Number of Factors for Different Time Spans 42

2.3 Survival Rates of the Factor Loadings 43

2.4 Survival Rates of the Factor Loadings under Different Restrictions 56

3.1 Summary statistics 97

3.2 Contemporaneous correlation coefficients of variables 97

CHAPTER 1

Macroeconomic Forecasting with Large Data Sets under Asymmetric Loss

1.1 Motivation

In forecasting macroeconomic series, the past decade has witnessed the increased availability and use of comprehensive data sets consisting of a large number of predictor time series. When forecasting macroeconomic aggregates like inflation or GDP, the appeal of such auxiliary data-rich sets is understandable: the additional informational content of the series helps improving forecasts compared to a benchmark (vector) autoregression of the variable to be predicted. At the same time, dealing with an increased number of predictor series poses problems, since the number of time observations is typically comparable with the number of series in such sets. This leads to imprecise coefficient estimates in an augmented predictive autoregression, and consequently to a trade-off between availability and usability of information. The literature has therefore focussed on complexity reduction and information extraction. Factor-based forecasting models, for which it is assumed that unobserved common components of the auxiliary series are good predictors for the variable of interest, are particularly popular in this respect.

Since the predictors are not observed directly for factor-based forecasts, the forecasting procedure boils down to estimating a feasible predictive regression using lags of the dependent variable and *extracted* factors as right-hand side variables. Several contributions have shown that a relatively small number of estimated factors successfully summarize the contemporaneous information in the data set of predictors. Stock and Watson (2002c) demonstrate Principal Component Analysis [PCA] of the predictors to produce consistent estimates of the space spanned by the common factors. Their factor model forecasts outperforms other benchmark models to forecast personal income and output growth; see also the earlier work in Stock and Watson (1998). Focussing on estimation and inference in approximate factor models, Bai (2003) derives asymptotic distributions and uniform convergence results while Bai and Ng (2002) provide information criteria for estimating the number of factors; see also Alessi et al. (2010).

The popularity of factor models in forecasting is reflected by the large number of contributions in the applied literature. Ludvigson and Ng (2009a,c) use factors from a large number of macroeconomic series to predict excess bond returns and to show that the predictability of future excess returns is related to macroeconomic activity. These are just the tip of the iceberg; see Marcellino et al. (2003), Artis et al. (2005), den Reijer (2005), Forni et al. (2005), Banerjee et al. (2008), Engel et al. (2012) or Godbout and Lombardi (2012) to name but a few more contributions to the literature on factor-based forecasting. While there are alternative approaches such as soft/hard thresholding or forecast combinations, they appear to be less popular than factor-based models. One reason to prefer factor-based forecasting procedures may be their interpretability; see e.g. the discussion in Ludvigson and Ng (2009a,c). For instance, Ludvigson and Ng (2009c) regress each macroeconomic variable in their data set on the PCA-extracted factors. The R^2s of these regressions are informative of the relations between the factors and the variables. They are thus able to identify e.g. stock market, inflation or real factors. More recently, Hacioglu Hoke and Tuzcuoglu (2014) work on factor augmented VAR models with a threshold structure of the loadings (which are dynamic in their setup). The periods where the loadings are set to zero or where the factors load more heavily on the variables are also informative on the relations between factors and variables. The point is that predictors with economic meaning prevent the interpretation of forecasting procedures as "crystal-ball" or "black-box" econometrics and are more likely to produce forecasts understandable by wider audiences.

The focus of the work cited above is on forecasts which are optimal in the mean squared-error [MSE] sense, i.e. on procedures minimizing the expected squared forecast error. The literature documents, however, a significant number of cases where more general – and in particular asymmetric – cost-of-error functions are employed. For instance, IMF and OECD forecasts of the deficit of G7 countries are found by Artis and Marcellino (2001) to be systematically biased towards over or under-prediction when compared with MSE-optimal forecasts. Elliott et al. (2005b) propose formal methods of inference on the degree of asymmetry of the loss function and testing the rationality of forecasts; see also Patton and Timmermann (2007b). Building on the work of Elliott et al., Christodoulakis and Mamatzakis (2008, 2009) find asymmetric preferences of EU institutional forecasts. Clements et al. (2007) discuss the loss function of the Federal Reserve and Capistrán (2008) even finds that, for inflation, the forecasting preferences of the Fed are time-varying. The loss function of the Bank of Canada is analyzed by Pierdzioch et al. (2011). More recently, Tsuchiya (2016) examines the asymmetry of the loss functions of the Japanese government, the IMF and private forecasters for Japanese growth and inflation forecasts.

We therefore study factor-augmented forecasting under asymmetric loss. For a given predictive model, there is little debate as to how to obtain point forecasts under a given loss function: it has been known since Weiss and Andersen (1984) and Weiss (1996) that the forecast model should be estimated under the relevant loss.[1] Estimation of the feasible predictive regression under the relevant loss would therefore improve forecasts. This prompts the question, first, whether such estimation may indeed be conducted with estimated factors in a manner analogous to the MSE-optimal case. Less obvious however, is the second question of whether the forecast model should be the same under any asymmetric loss function. To put it bluntly, are the PCA-extracted factors still forecast-relevant under an asymmetric loss function? Considering the theory of forecasting under asymmetric loss functions, see Granger (1969), Granger (1999), Weiss (1996), Christoffersen and Diebold (1996), McCullough (2000), Elliott and Timmermann (2004), Elliott et al. (2005b), Patton and Timmermann (2007a) or Patton and Timmermann (2007b), the least what may be expected is that the relative importance, as a predictor changes, for the extracted factors or even for the lags of the dependent variable in the augmented predictive autoregression. So, rather than relying on the summarizing power of, say, the first principal component, one may have to select the predictors (lagged dependent variables or factors) that are most informative under the relevant loss.[2] Third, perhaps even more importantly, one should ask whether the usual factor extraction does actually capture all information relevant under the given loss function. PCA essentially delivers linear combinations of the "many predictors" data set. In a linear predictive model under squared-error loss, this may be a convenient dimensionality reduction procedure. But the optimal forecast function under an asymmetric loss function may depend on the auxiliary series in a non-linear fashion, even if the optimal forecast function is linear in the MSE-optimal case. Thus, the informational content of the data set may not be fully exploited under an asymmetric loss function.

Our contributions are as follows. We show in Section 1.2 that, regularity conditions provided, one may indeed use PCA-extracted factors as predictors even when estimating forecast regressions using the relevant loss function. To make sure that relevant information is not wasted, we make use in Section 1.3 of the insight that the optimal point forecast under a general loss depends on the conditional variance of the variable to be predicted (Christoffersen and Diebold, 1996; Patton and Timmermann, 2007b). Thus, adding information on the *volatility* of the series to be predicted in the forecasting model improves forecasts under asymmetric loss. While the

[1] An alternative, more demanding, procedure is to model the entire predictive distribution and derive the point forecasts based on it; see e.g. McCullough (2000) for an ingenious bootstrap-based version.

[2] In fact, focussing on extracting the factors with the highest associated eigenvalue might not be a good idea in the MSE-case either, since a factor even if explaining most of the variance of the raw predictor series, need not capture the information relevant for forecasting.

volatility of interest is not observed directly, it is plausibly related to the variability of the auxiliary series. The relation is not a forced one, since the volatility of the overall economic environment should be reflected – at least to some extent – by the volatility of all series involved. This common component can in turn be extracted from the auxiliary data set. Concretely, we extract additional factors from the log-squared residuals of the factor model to increase the quality of the forecasts under the relevant loss. This delivers a larger number of predictors, of which not all need be equally relevant. To find the ones with the highest predictive power, we resort to a suitable information criteria.[3]

We then illustrate the proposed procedure in Section 1.4 by means of a forecasting exercise with US personal income, industrial production, unemployment rate and retail sales. We use a data set which has become widely known as the "Stock and Watson" data set (Stock and Watson, 2005). Expanded by Ludvigson and Ng (2009a), the data set spans the period from January 1964 to the end of 2007. Although the original data set includes more time series than we work with here, the use of this particular selection of 131 series has been quite popular in the literature; see e.g. Belviso and Milani (2006), Boivin and Ng (2006), D'Agostino and Giannone (2006), Ludvigson and Ng (2009a,c) and Bai and Ng (2011). The detailed description and other features of the data can be found in Appendix 1.D. Here, we are interested in one-year-ahead forecasts; working with monthly data, we thus work at forecast horizon $h = 12$. We compare the average forecast losses of all four variables in every single case we look into. We find, expectedly, that average losses of forecasts produced under the relevant loss function give smaller losses compared to the losses produced by forecasts obtained via OLS estimation of the predictive regression. At the same time, we also show that adding information from the volatility of the series and having parsimonious models by assessing the relevance of the extracted factors improve the average losses.

The final section concludes, and some mathematical details and additional results have been gathered in the Appendix.

1.2 The basic forecasting problem

Let y_t be the series for which an h-step ahead forecast is required. Given the available information set $\mathcal{F}_t = \{f_{t,k}, y_t, y_{t-1}, \ldots\}$, the optimal forecast is given by

$$y_{t+h}^{opt} = \arg\min_{y_{t+h}^*} \mathrm{E}\left(\mathcal{L}\left(y_{t+h} - y_{t+h}^*\right) | \mathcal{F}_t\right), \tag{1.1}$$

[3]The issue of model selection is not restricted to our setup: e.g. Schumacher (2007) compares the forecast accuracy of variety of factor models to MSE-predict German GDP, and finds that results may change when different information criteria to select factors are used.

where $\mathcal{L}(\cdot)$ is the relevant loss function quantifying the cost incurred by discrepancies between a given forecast y^*_{t+h} of the variable y at some time $t+h$ and the actual realization y_{t+h}. According to Granger (1999), loss functions should be uniquely minimized at the origin, and be quasi-convex. We shall work with a specific class of loss functions, introduced by Elliott et al. (2005b); a forecast y^*_{t+h} is thus evaluated by means of

$$\mathcal{L}\left(y_{t+h} - y^*_{t+h}\right) = \left(\alpha + (1-2\alpha)\,\mathbb{I}\left(y_{t+h} - y^*_{t+h} < 0\right)\right) \left|y_{t+h} - y^*_{t+h}\right|^p . \tag{1.2}$$

This class of loss functions is quite flexible: it includes as special cases the widely used symmetric (for $\alpha = 0.5$) and asymmetric (for $0 < \alpha < 0.5$ or $0.5 < \alpha < 1$); linear and quadratic loss functions (for $p = 1$ and $p = 2$). Moreover, it only requires mild moment conditions on y_t, in contrast e.g. to the well-known linex loss.

We start with the usual linear forecasting model

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k f_{t,k} + v_{t+h}, \quad t = 1, 2, \ldots, T\,, \tag{1.3}$$

where the forecast error v_{t+h} cannot be predicted under $\mathcal{L}$. This does not imply, however, that v_{t+h} could not be forecast under another loss function. The lack of predictability of v_{t+h} under $\mathcal{L}$ implies that the so-called generalised forecast error $\mathcal{L}'\left(v_{t+h}\right)$ is uncorrelated with the predictors y_{t-j+1} and $f_{t,k}$; see Granger (1999) and Patton and Timmermann (2007a). The optimal forecast is thus given by

$$y^{opt}_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k f_{t,k}. \tag{1.4}$$

In practice, one resorts to a two-stage procedure, given that observations on N auxiliary variables $x_{t,i}$ are available, from which $f_{t,k}$ may be estimated in a first stage. Maintaining the typical assumption of linear measurement equations for the factors, we have that

$$x_{t,i} = \sum_{k=1}^{r} \lambda_{i,k} f_{t,k} + u_{t,i}. \tag{1.5}$$

With additional conditions on $\lambda_{i,k}$ and $u_{t,i}$ (in particular orthogonality of the common and idiosyncratic components $f_{t,k}$ and $u_{t,i}$), extraction of the unknown factors can be conducted, leading to $\hat{f}_{t,k}$ (we resort to PCA to this end). This ultimately takes us to the feasible predictive regression

$$y_{t+h} = c + \sum_{j=1}^{q} a_j y_{t-j+1} + \sum_{k=1}^{r} b_k \hat{f}_{t,k} + v_{t+h}, \tag{1.6}$$

to be estimated under the relevant loss in a second stage, i.e.

$$\tilde{c}, \tilde{a}_j, \tilde{b}_k = \underset{c^*, a_j^*, b_k^*}{\arg\min} \frac{1}{T} \sum_{t=q}^{T-h} \mathcal{L}\left(y_{t+h} - c^* - \sum_{j=1}^{q} a_j^* y_{t-j+1} - \sum_{k=1}^{r} b_k^* \hat{f}_{t,k} \right), \tag{1.7}$$

from which the forecast is obtained as

$$\tilde{y}_{t+h}^{opt} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k}. \tag{1.8}$$

Its quality hinges on the precision of the factor approximation; recall that factors cannot be consistently estimated in a fixed-N setup.

The justification to use the feasible forecast from (1.8) is provided by the following proposition establishing its consistency as $T, N \to \infty$ for the unfeasible optimal forecast from (1.4) under the relevant loss $\mathcal{L}$.

Proposition 1. *Let the auxiliary variables $x_{t,i}$ obey Assumptions A-E in Bai (2003). Furthermore, assume that the factors $f_{t,k}$ and the forecast errors v_{t+h} are strictly stationary and ergodic, and that the generalised forecast errors $\mathcal{L}'(v_{t+h})$ satisfy $\mathrm{E}\left(\mathcal{L}'(v_{t+h}) | y_t, y_{t-1}, \ldots, f_{t,k}\right) = 0$ and have no atom at 0. Finally, let all series have finite moments of order p with p from (1.2) integer and positive. It then holds for the estimated optimal forecast from (1.8) that, point-wise in t,*

$$\tilde{y}_{t+h}^{opt} \xrightarrow{p} y_{t+h}^{opt}$$

as $N, T \to \infty$ such that $T/N \to 0$.

Proof: *See Appendix 1.B.*

Remark 2. *Assumptions A-E in Bai (2003) ensure the uniform (in t) consistency of the extracted factors, which, in that framework, may be heteroskedastic and even locally trending. The additionally required strict stationarity simplifies the proofs; while it is slightly more restrictive than the often made assumption of weak stationarity (see e.g. Stock and Watson, 2002c), it is a convenient price to pay for being able to use non-MSE loss functions. Strict stationarity of the factors might be relaxed at the expense of additional conditions, but we do not pursue the topic here as we would rather focus on the forecasting procedure than on more involved technical details. The critical requirement is that the generalised forecast error is a martingale difference sequence, which is a standard condition in the literature on forecasting under asymmetric loss (Patton and Timmermann, 2007a). In a nutshell, the forecast errors must be unforecastable* *under the relevant loss. The finiteness of the pth order moments ensures that the forecast risk is finite and an optimal forecast exists.*

Remark 3. *In factor models, the factors are only identified up to a rotation. But it follows from the proof that rotations do not affect the result: essentially,* $\sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k}$ *consistently estimates* $\sum_{k=1}^{r} b_k f_{t,k}$ *which is the quantity required for forecasting* y_{t+h}*. E.g. Bai and Ng (2006) consider this explicitly; to keep notational effort at a minimum, we assume identification directly.*

Remark 4. *The loss function does not play any role in estimating the factors, but only in the subsequent forecasting step. The main reason to do so is to maintain the interpretability of the factors as economic driving forces (not depending on individual loss preferences), but we also wish to stay in line with the literature on factor-based forecasting. While Tran et al. (2014) discuss estimation of factors under asymmetric linear and asymmetric quadratic losses, these losses refer to the idiosyncratic components and not to the actual forecast errors; we leave the integration of the two approaches to further work.*

1.3 Extracting additional relevant information

The two-step procedure for forecasting under asymmetric loss discussed in the previous section is the natural extension of the original method of Stock and Watson (2002c), for which the second step – i.e. estimation of the predictive regression – has been modified to account for the use of a specific loss function. But we should ask at this point whether the first stage – i.e. extracting the information carried by the auxiliary variables $x_{t,i}$ – is to be left unmodified. In other words, is the factor model (1.5) exhausting the possibilities of finding predictors for y_{t+h} under the relevant loss?

It should be pointed out that the linear model (1.5) is only sufficient under conditions which are not plausible for macroeconomic data sets. Namely, Patton and Timmermann (2007b) show that, for loss functions of the type given in (1.2), the optimal forecast has the form

$$y_{t+h}^{opt} = \mathrm{E}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i}\right) + C\sqrt{\mathrm{Var}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i}\right)} \tag{1.9}$$

for some constant C depending on the loss function and the shape of the conditional distribution.[4] The first summand on the r.h.s. of (1.9) is nothing else that the conditional mean which the original factor-based model does indeed capture. The coefficient C, and thus the second

[4]Their result actually holds for any homogenous loss function.

summand, is zero e.g. when $\alpha = 0.5$ and $p = 2$, or when $\alpha = 0.5$ and the conditional distribution of y_{t+h} is symmetric, but not in general. When estimated under the relevant loss, the intercept c of the predictive regression (1.3) only captures the *average* of the so-called bias term $C\sqrt{\mathrm{Var}\left(y_{t+h} \middle| y_t, y_{t-1}, \ldots, x_{t,i}\right)}$ and misses the fact that the conditional standard deviation of y_{t+h}, *if time-varying*, is actually a predictor for y_{t+h} under $\mathcal{L}$.

And indeed, the volatility of macroeconomic variables is not constant in general. The Great Moderation is the perhaps best known case of time-varying volatility. The term coins the downward trend in the variance of inflation and economic growth since the 1980s (e.g., Stock and Watson, 2002b); Clark (2009) finds that the recent financial crisis has reversed the trend, thus strengthening the evidence of time-varying volatility. Along the same lines, Sensier and van Dijk (2004) find that four out of five of over two hundred U.S. macroeconomic time series exhibit unconditional volatility changes during the period 1959-1999.

What is more, it is expected that such volatility trends are common to the variables in the data set used for forecasting: the series stem, after all, from the same economic environment. Thus, we may resort to *the same* data set $\{x_{t,i}\}$ in order forecast the conditional standard deviation of y_{t+h}.

To exploit the above insight, we assume a stochastic volatility model of the form

$$v_{t+h} = e_t \, e^{\frac{1}{2}\left(g_t + \sum_{l=1}^{s} \xi_l h_{t,l}\right)}.$$

We follow Nelson (1991) in using the exponential "link" function, since it allows us to avoid positivity restrictions on the components g_t and $h_{t,l}$ and assume – in line with the very idea of factor-based forecasting – that $h_{t,l}$ could be forecast using information from the auxiliary series $x_{t,i}$; g_t is an unforecastable component. When the conditional variance of the idiosyncratic components in the factor model depend in a similar manner on $h_{t,l}$, we write

$$u_{t,i} = e_{t,i} \, e^{\frac{1}{2}\left(g_{t,i} + \sum_{l=1}^{s} h_{t,l} \xi_{l,i}\right)},$$

where $g_{t,i}$ are individual volatility components specific for $x_{t,i}$. As usually, e_t and $e_{t,i}$ are standardised variables, mutually independent and independent of $h_{t,l}$, g_t and $g_{t,i}$. Then,

$$\log u_{t,i}^2 = \log e_{t,i}^2 + g_{t,i} + \sum_{l=1}^{s} \xi_{l,i} h_{t,l},$$

which is nothing else than a factor model for the log squares of $u_{t,i}$ with $h_{t,l}$ the common components and $\log e_{t,i}^2 + g_{t,i}$ the idiosyncratic ones.

Since the variables $u_{t,i}$ are not observed directly, we resort to the idiosyncratic components extracted in the first-stage PCA. Thus we are now able to extract $h_{t,l}$ from $\log \hat{u}_{t,i}^2$ using a second-stage PCA, leading to $\hat{h}_{t,l}$. Note that the factors $f_{t,k}$ themselves may be (conditionally) heteroskedastic; we assume that they do not bear additional predictive power for the conditional variance of y_{t+h}, but one may of course consider their log squares when extracting $h_{t,l}$.

This is related to decomposition of the yield spreads in Ludvigson and Ng (2009a,c). In both papers, additional information carried by the yield risk premium (or term premium) is acknowledged, due to the inability of the yield curve to explain business cycle variations in bond risk premia. The yield risk premium can be seen as an idiosyncratic error which should be constant under the expectation hypothesis. Ludvigson and Ng estimate this term via the average multi-step estimates of bond returns. They show that the predictive factors are not sufficient to display the countercyclical form of bond risk premia since the predictive power of these factors does not imply explaining the yield curve. In this respect, the additional information used, namely the yield risk premium, parallels the volatility factor we use in this paper.

Equation (1.9) shows that a nonlinear forecast may be better suited in an asymmetric loss context. Clearly, extracting factors from $\log u_{t,i}^2$ is not the only way to consider nonlinearities; for instance Bai and Ng (2008a) employ quadratic PCA. But Equation (1.9) motivates us to look directly for variables driving the volatility.

Ideally, we would include a term of the form $Ce^{\frac{1}{2}\sum_{l=1}^{s}\xi_l\hat{h}_{t,l}}$ in the predictive regression with additional parameters ξ_l (with g_t not being predictable, $e^{1/2g_t}$ is absorbed in the error component e_t multiplicatively). But a non-linear regression equation is perhaps too cumbersome to deal with numerically, even if we must anyway resort to numerical optimization under non-MSE loss.[5] We therefore linearize the exponential, $e^x \approx 1 + x$, and trade some misspecification in exchange for increased clarity of the final procedure.

The component g_t is in principle not forecastable, at least not from $x_{t,i}$, and we treat it as such by absorbing it in the forecast error. We thus obtain as estimated predictor for y_{t+h}

$$\tilde{y}_{t+h}^{opt} = \tilde{c} + \sum_{j=1}^{q}\tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r}\tilde{b}_k\hat{f}_{t,k} + \sum_{l=1}^{s}\tilde{\xi}_l\hat{h}_{t,l}, \tag{1.10}$$

where the parameter estimates are obtained like before by minimising the observed forecast loss.

Due to the linearization, the estimators $\tilde{\xi}_l$ in (1.10) do not converge to the population values. The following proposition guarantees that the fitted predictor is the best *linear* predictor under the given loss.

Proposition 5. *Define the (unfeasible) linear predictor*

$$\pi(y_t, f_t, h_t) = c^* + \sum_{j=1}^{q}a_j^* y_{t-j+1} + \sum_{k=1}^{r}b_k^* f_{t,k} + \sum_{l=1}^{s}\xi_l^* h_{t,l}$$

[5] See Demetrescu (2006) for a tailored optimization method.

and assume that $\sup_t \left|\hat{h}_{t,l} - h_{t,l}\right| = o_p(1)$*. Under the assumptions of Proposition 1, it holds for* $\tilde{y}^{opt}_{t+h}$ *from* (1.10) *that*

$$\tilde{y}^{opt}_{t+h} \xrightarrow{p} \underset{c^*, a^*_j, b^*_k, \xi^*_l}{\arg\min} \; \mathrm{E}\left(\mathcal{L}\left(y_{t+h} - \pi(y_t, f_t, h_t)\right)\right)$$

pointwise in t*.*

Proof: *Analogous to the proof of Proposition 1 and omitted.*

Remark 6. *In the case of the squared-error loss, the bias-variance decomposition of the MSE indicates that the fitted linear model minimizes the expected squared difference between the linear fit and the nonlinear regression curve (where the expectation is taken with respect to the marginal distribution of the predictors). In the case of asymmetric power losses, such a clean decomposition is not available, but the interpretation of the proposition remains the same.*

Remark 7. *The quality of the linear approximation depends on the signal-to-noise ratio in the series* $\sum_{l=1}^{s} \xi_l \hat{h}_{t,l}$*. One could improve it by taking a quadratic approximation for the exponential,* $\mathrm{e}^x \approx 1 + x + x^2/2$*. When not imposing the coefficient restrictions resulting from the quadratic approximation of the exponential function to avoid further numerical complications, this results in a linear model with interactions,*

$$\tilde{y}^{opt}_{t+h} = \tilde{c} + \sum_{j=1}^{q} \tilde{a}_j y_{t-j+1} + \sum_{k=1}^{r} \tilde{b}_k \hat{f}_{t,k} + \sum_{l=1}^{s} \sum_{m=1}^{s} \tilde{\xi}_l \tilde{\xi}_m \hat{h}_{t,l} \hat{h}_{t,m}.$$

To sum up, the factor-based forecasting procedure is modified under asymmetric loss as follows.

1. Clean/prepare the auxiliary data set and the variable to be predicted.
2. Extract factors from auxiliary series (PCA).
3. Extract factors (demean, standardise, PCA) from log-squared extracted idiosyncratic components.
4. Augment the predictive autoregression with the factors extracted in steps 2 and 3.
5. Estimate under the relevant loss.
6. Suitably select the predictors to enter the predictive model.

Compared to the usual factor-based forecasting approach, steps 3 and 5 are new and specific to forecasting under a general loss function. Step 6 should of course be conducted even under squared-error loss, but requires here a careful consideration of the used selection tool. Concretely, to conduct predictor selection in (1.10), we resort to an information criterion, but tailored to the relevant loss. For a model of complexity k, we thus compute

$$AIC_{\mathcal{L}}\left(k\right)=\frac{2}{p}\ln\left(\sum\mathcal{L}\left(\hat{v}_{t+h}(k)\right)\right)+\frac{2k}{T}$$

with $\hat{v}_{t+h}(k)$ in-sample fitted errors from the respective model, and choose the model minimizing the criterion. See Appendix 1.A for a justification of this particular choice.

We work with an information criterion because of the widespread use of information criteria in general, but partly also for computational convenience; we also examined the numerically more involved least absolute shrinkage and selection operator [LASSO] (Tibshirani, 1994) as an alternative, alongside with refinements due to Belloni and Chernozhukov (2013). Other choices such as targeting the predictors á la Bai and Ng (2008a) (see also Dias et al., 2010) are not considered, but may of course be incorporated in the forecasting procedure.

We present in the following section the empirical results obtained using only the tailored information criterion AIC$_{\mathcal{L}}$ for model selection. The corresponding LASSO and post-fit LASSO results are presented in Appendix 1.C. While we find that they (in particular the post-fit LASSO) improve on AIC$_{\mathcal{L}}$, the computational requirements are higher and we leave the decision of which model selection procedure to use to the practitioner.

1.4 Forecasting under asymmetric loss

The goal of the exercise is to forecast several macroeconomic variables, such as Personal Income (PI), Industrial Production (IP), Unemployment Rate (UN) and Retail Sales (SL), under asymmetric loss. We evaluate the out-of-sample forecasts that use the factors recursively extracted from the auxiliary data. The factors are extracted by PCA analysis in a *linear* fashion. We pursue the empirical analysis by taking them as observable. The exercise follows that of Ludvigson and Ng (2009c)'s.

1.4.1 Setup

The data set employed for the forecasting exercise is often referred to as the Stock and Watson data set (Stock and Watson, 2005) which consists of 131 macroeconomic aggregates. Ludvigson and Ng (2009c) updated this data set so it now spans the time period 1964:01 – 2007:12. The consistency of the estimated forecast function relies, among others, on the assumption that observable series are stationary. The series are therefore transformed to stationarity by taking differences, by taking logarithms – and in some cases by doing both; see Appendix 1.D for details. Finally, all transformed variables are standardized to have zero sample mean and unit sample variance for factor extraction.

We use a recursive pseudo out-of-sample forecasting scheme to allow for a comparison of the different forecasting procedures considered in the following. Concretely, we start with data from 1964:1 through 1984:12; we run the forecasting regression with dependent variables from 1965:1 to 1984:12 and predictors from 1964:1 to 1983:12. The outcome is used to forecast PI, IP, UN and SL for 1985:12. We then expand the data set by one period to obtain the forecasts for 1986:1. The procedure is iterated until we obtain the last forecast, for 2007:12. (At the last step, the independent variables from 1964:1 through 2005:12 and dependent variables from 1965:1 to 2006:12 are used to run the forecasting regressions to forecast 2007:12.)

1.4.2 Extracted factors

This section directs our focus to forecasting by using PCA-extracted factors from the data. The results of this section shed light on whether forecasting with factors under asymmetric loss is effective. Furthermore, we emphasize the importance of model selection and additional information presented by volatility factor(s).

One of the common issues associated with factor-based forecasting approaches is the number of factors to be extracted from the auxiliary data set. To set this in stone, we start by performing the information criteria developed by Bai and Ng (2002), and used by Ludvigson and Ng (2009a,c) and Bai and Ng (2011).[6] The criteria find eight factors in the Stock and Watson data set. Factors are identified up to a rotation, so a comprehensive interpretation of extracted factors is not straightforward. Stock and Watson (2002a) and Ludvigson and Ng (2009c) report marginal R^2s of the regressions of each of the series against each of the eight factors they infer from the information criteria. In line with their pre-classification of the dataset, they relate these factors with real economy, output and unemployment series, Treasury Bills, commodity prices and such. Note, however, that the forecasting procedure does not hinge on this classification.

For a closer look on the number of factors, we employ the tailored $\text{AIC}_{\mathcal{L}}$ for a preliminary check of the number of factors for the *full time span*. This preliminary exercise starts with selecting among the 8 largest PCA-extracted factors which are chosen by the Bai and Ng (2002) information criteria. In the second step, 9 factors are extracted from the auxiliary data and selection is conducted among these 9, and so on. We stop at selection among the 15 largest

[6] Bai and Ng (2002) information criteria do *not* consider generalised loss functions. We apply these criteria to give a preliminary idea about the number of the factors.

PCA-extracted factors. The factors in each step are used in the predictive regressions to forecast all four variables of interest after being subject to the model selection. The potential forecast relevance of the factors then assessed; Table 1.1 reports the model (i.e. the factors) chosen by minimizing AIC$_{\mathcal{L}}$ among all factors in that particular step. The loss function is asymmetric quadratic with $p = 2$ and $\alpha \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$.

As shown in the columns of Table 1.1, not all factors in each step are selected as predictors, at least for the full data span. For example, for forecasting PI, in case of $\alpha = 0.1$, all but fourth factor are selected when selecting among the first 8 factors in total. For the same variable, when $\alpha = 0.5$, the forth and sixth factors are not identified as forecast relevant in the first step. For all α, it turns out that the first 8 factors given by the information criteria are not all forecast-relevant.[7] Increasing the number of factors to select from one by one, the already selected factors do not generally change. In the last step of our exercise, we contemplate all 15 PCA-extracted factors and note that some of the additional ones appear to be forecast relevant, while some of the commonly used 8 largest factors do not. Changes on the selected factors are observed depending on the chosen α values. This emphasises the differences on the relative importance of the factors which changes with the loss function.

Evidence from this preliminary exercise suggests that the 9^{th} factor is rarely chosen by the tailored AIC$_{\mathcal{L}}$ while forecasting PI and PI. On the contrary, it is consistently chosen in the cases of UN and SL, for all α values. Additionally, factors beyond 9 appear to be forecast relevant. Thus, we use 15 factors (the largest PCA-extracted ones) as benchmark rather than 8 largest factors found by the information criteria. To keep the complexity tractable, we do not consider classical factors beyond these.

The extracted volatility factor(s) give(s) information which is not (linearly) contained in the original series. According to the mentioned information criteria,[8] the PCA of the log-squared residuals from the first-step factor analysis leads to only one additional factor to be taken into account. We consider it as a predictor along with the factors extracted from the data in the first step. While selecting the concrete predictive model for a given span of observations, the volatility factor is subject to model selection with the tailored AIC alongside the other factors. We also consider the squared volatility factor to better account for nonlinearities.

[7]The objective function of the AIC$_{\mathcal{L}}$ targets the dependent variable whereas the PCA analysis aims to maximize the variance explained by factors. Due to the difference in the objective functions, the factors selected by the information criteria do not always appear to be forecast relevant.

[8]Following Bai and Ng (2002), we rely on PC_{p2} and PC_{p3} as the other criteria tend to – unrealistically – over-parameterize the model in our case.

Table 1.1. Factors selected for predicting for all predictor series by $AIC_{\mathcal{L}}$; full data span

$\alpha = 0.1$

	8	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Personal Income	1, 2, 3, 5, 6, 7, 8	1, 2, 3, 5, 6, 7, 8	1, 3, 5, 6, 7, 8, 10	1, 2, 5, 6, 7, 8, 10, 11	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15
Industrial Production	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 7, 8, 9, 10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15
Unemployment Rate	1, 2, 3, 4, 6, 7, 8	1, 2, 3, 4, 6, 7, 8, 9	1, 2, 7, 8, 9, 10	1, 2, 4, 6, 7, 8, 9, 10, 11	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14
Retail Sales	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8, 9	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

$\alpha = 0.3$

	8	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Personal Income	1, 2, 3, 5, 6, 7, 8	1, 2, 3, 5, 6, 7, 8	1, 2, 3, 5, 6, 7, 8, 10	1, 2, 3, 5, 6, 7, 8, 10, 11	1, 2, 3, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15
Industrial Production	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8, 10	1, 2, 3, 4, 5, 6, 7, 8, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 15
Unemployment Rate	1, 2, 3, 6, 7, 8	1, 2, 3, 6, 7, 8, 9	1, 2, 3, 6, 7, 8, 9, 10	1, 2, 3, 6, 7, 8, 9, 10, 11	1, 2, 3, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Retail Sales	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8, 9	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

$\alpha = 0.5$

	8	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Personal Income	1, 2, 3, 5, 7, 8	1, 2, 3, 5, 7, 8	1, 2, 3, 5, 7, 8, 10	1, 2, 3, 5, 6, 7, 8, 10, 11	1, 2, 3, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15
Industrial Production	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8, 10	1, 2, 3, 4, 5, 6, 7, 8, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14
Unemployment Rate	1, 2, 3, 6, 7, 8	1, 2, 3, 6, 7, 8, 9	1, 2, 3, 6, 7, 8, 9, 10	1, 2, 3, 6, 7, 8, 9, 10, 11	1, 2, 3, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Retail Sales	1, 2, 4, 5, 6, 7, 8	1, 2, 4, 5, 6, 7, 8, 9	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

$\alpha = 0.7$

	8	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Personal Income	1, 2, 3, 5, 7, 8	1, 2, 3, 5, 7, 8	1, 2, 3, 5, 7, 8, 10	1, 2, 3, 5, 7, 8, 10, 11	1, 2, 3, 5, 7, 8, 10, 11, 12	1, 2, 3, 5, 7, 8, 10, 11, 12, 13	1, 2, 3, 5, 7, 8, 10, 11, 12, 13, 14	1, 2, 3, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15
Industrial Production	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8, 10	1, 2, 3, 4, 5, 6, 7, 8, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15
Unemployment Rate	1, 2, 3, 6, 7, 8	1, 2, 3, 6, 7, 8, 9	1, 2, 3, 6, 7, 8, 9, 10	1, 2, 3, 6, 7, 8, 9, 10, 11	1, 2, 3, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Retail Sales	1, 2, 4, 5, 6, 7, 8	1, 2, 4, 5, 6, 7, 8, 9	1, 2, 4, 5, 6, 7, 8, 9, 10	1, 2, 4, 5, 6, 7, 8, 9, 10	1, 2, 4, 5, 6, 7, 8, 9, 10, 12	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 13	1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14	1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15

$\alpha = 0.9$

	8	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Personal Income	1, 2, 3, 7, 8	1, 2, 3, 7, 8	1, 2, 3, 5, 7, 8, 10	1, 2, 3, 4, 5, 7, 8, 10, 11	1, 2, 3, 4, 5, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13	1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14	1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15
Industrial Production	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8	1, 2, 3, 4, 5, 6, 7, 8, 10, 11	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 12, 15
Unemployment Rate	1, 2, 3, 6, 7, 8	1, 3, 6, 7, 8, 9	1, 2, 3, 6, 7, 8, 9, 10	1, 2, 3, 6, 7, 8, 9, 10, 11	1, 2, 3, 6, 7, 8, 9, 10, 11, 12	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
Retail Sales	1, 2, 4, 5, 6, 7, 8	1, 2, 4, 5, 6, 7, 8, 9	1, 2, 4, 5, 6, 7, 8, 9, 10	1, 2, 4, 5, 6, 7, 8, 9, 10	1, 2, 4, 5, 6, 7, 8, 9, 10, 12	1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13	1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14	1, 2, 4, 5, 6, 7, 8, 9, 10, 12, 13, 14, 15

Notes: Asymmetric quadratic loss; see the text for details. The number of factors considered given a particular α and variable are given in bold. The analysis is for the whole time span.

1.4.3 Results

In this section, we discuss the results when model selection is conducted with the tailored information criterion $\text{AIC}_{\mathcal{L}}$. For each α, we first estimate the respective predictive regression by ordinary least squares (OLS) relying on the regressors in the benchmark models in recursive manner. We construct one-year-ahead forecasts in each given step and evaluate the occurring loss via the forecast errors under the relevant loss function. This approach is henceforth named as OLS-Asymmetric Loss ($OLS-AL$). The second route to take here is estimating the regression coefficients numerically directly by the aggregated observed loss and using them to construct forecasts. This approach is named as Asymmetric Loss (AL) henceforth. For each variable of interest, first $OLS - AL$ losses are presented and followed by the AL losses. For $\alpha = 0.5$, the results of consecutive columns are the same, since for $p = 2$ and $\alpha = 0.5$ the quadratic loss is recovered.

Concentrating on the evaluation of the forecasts obtained using OLS vs. those obtained via estimation under the relevant loss, one expects the average forecast loss of AL to be smaller than the loss which occurs under $OLS - AL$; see the early work of Weiss and Andersen (1984).

We consider six cases in total. The first case uses only 15 factors for the forecasting exercise. The second case also includes the factor extracted from the log-squared idiosyncratic components $\widehat{u}_{t,i}$. Thus, there are in total 16 factors for this case. The third case adds the squared volatility factor after which we end up with 17 factors. We do not conduct model selection for Case 1, Case 2 or Case 3. Case 4 is the counterparty of Case 1 with model selection by the tailored $\text{AIC}_{\mathcal{L}}$. Similarly, Case 5 and 6 are model selection versions of Case 2 and 3, respectively. Note that the model selection is performed in each recursive step. Moreover, one lag of the dependent variable is added to the set of predictors in all cases (and is subject to model selection in Cases 4, 5 and 6). A second lag did not improve forecasting ability in any of the cases or for any of the loss functions so we do not present those results here. Appendix 1.C contains the additional results based on LASSO and post-fit LASSO model selection and estimation.

Table 1.2 summarizes the pseudo out-of-sample average forecast loss for each of the six cases. We fix $p = 2$ and allow for different degrees of asymmetry by considering five αs for the loss function in Equation (1.2). For each of the six cases we consider, the goal is to forecast PI, IP, UN and SL under two alternatives of forecast evaluation.

Evaluating the forecasts by the asymmetric loss function of choice leads to lower average losses with the only exception of forecasting IP with $\alpha = 0.7$ in Case 5.[9]

In some cases, we see that adding one extra factor, the volatility factor, improves the forecast accuracy. $OLS - AL$ and AL losses reported in case of $\alpha = 0.1$ numerically demonstrate the

[9] Model selection conducted by LASSO and post-fit LASSO leads to similar findings; see Appendix 1.C.

Table 1.2. Losses Evaluated for OLS-Asymmetric Loss and Asymmetric Loss

Alpha	Cases	PI_{OLS-AL}	PI_{AL}	IP_{OLS-AL}	IP_{AL}	UN_{OLS-AL}	UN_{AL}	SL_{OLS-AL}	SL_{AL}
0.1	Case 1	0.2536	0.2023	0.1793	0.0988	0.0092	0.0054	0.8751	0.6065
	Case 2	0.2532	0.2004	0.1845	0.1019	0.0092	0.0054	0.8978	0.6089
	Case 3	0.2520	0.2042	0.1785	0.1009	0.0093	0.0055	0.8896	0.6129
	Case 4	0.2493	0.1866	0.1852	0.0913	0.0086	0.0054	0.8236	0.6023
	Case 5	0.2502	0.1872	0.1853	0.0911	0.0086	0.0054	0.8236	0.6023
	Case 6	0.2501	0.1871	0.1832	0.0935	0.0086	0.0054	0.8236	0.6023
0.3	Case 1	0.2369	0.2232	0.1695	0.1494	0.0098	0.0092	0.8668	0.8149
	Case 2	0.2365	0.2218	0.1728	0.1532	0.0099	0.0093	0.8794	0.8158
	Case 3	0.2366	0.2231	0.1695	0.1520	0.0100	0.0093	0.8805	0.8220
	Case 4	0.2345	0.2179	0.1650	0.1369	0.0094	0.0089	0.8235	0.7895
	Case 5	0.2345	0.2179	0.1650	0.1369	0.0094	0.0089	0.8235	0.7895
	Case 6	0.2345	0.2179	0.1655	0.1372	0.0094	0.0089	0.8235	0.7895
0.5	Case 1	0.2201	0.2201	0.1598	0.1598	0.0105	0.0105	0.8586	0.8586
	Case 2	0.2198	0.2198	0.1611	0.1611	0.0106	0.0106	0.8611	0.8611
	Case 3	0.2212	0.2212	0.1604	0.1604	0.0107	0.0107	0.8714	0.8714
	Case 4	0.2162	0.2162	0.1419	0.1419	0.0099	0.0099	0.7930	0.7930
	Case 5	0.2162	0.2162	0.1419	0.1419	0.0099	0.0099	0.7930	0.7930
	Case 6	0.2162	0.2162	0.1419	0.1419	0.0099	0.0099	0.7934	0.7934
0.7	Case 1	0.2034	0.2006	0.1500	0.1419	0.0111	0.0098	0.8503	0.7890
	Case 2	0.2031	0.2016	0.1494	0.1403	0.0113	0.0099	0.8427	0.7948
	Case 3	0.2058	0.2035	0.1514	0.1404	0.0114	0.0100	0.8623	0.8086
	Case 4	0.1980	0.1968	0.1288	0.1286	0.0106	0.0093	0.8033	0.7355
	Case 5	0.1980	0.1968	0.1289	0.1292	0.0106	0.0093	0.8034	0.7389
	Case 6	0.1980	0.1968	0.1293	0.1286	0.0106	0.0094	0.8033	0.7387
0.9	Case 1	0.1866	0.1537	0.1403	0.0848	0.0118	0.0064	0.8420	0.5554
	Case 2	0.1864	0.1555	0.1377	0.0834	0.0121	0.0066	0.8244	0.5674
	Case 3	0.1903	0.1577	0.1423	0.0839	0.0121	0.0065	0.8531	0.5840
	Case 4	0.1936	0.1484	0.1194	0.0819	0.0114	0.0063	0.8122	0.5309
	Case 5	0.1936	0.1484	0.1193	0.0835	0.0116	0.0064	0.8091	0.5357
	Case 6	0.1936	0.1484	0.1224	0.0816	0.0114	0.0065	0.8109	0.5356

Notes: The losses are evaluated using asymmetric quadratic loss functions within a recursive pseudo-out-of-sample setup. See the text for details. The dataset for factor extraction includes the dependent variables.

improvement from Case 1 to Case 2. Due to the exceptions, this cannot be generalised over all cases and all variables. Adding the squared volatility factor improves the forecasts of some variables under different loss forecast asymmetries, such as for PI_{OLS-AL} for $\alpha = 0.1$. However, for the same asymmetry switching from Case 2 to Case 3 in AL, PI forecast losses point out otherwise, illustrated by increasing forecast losses with the inclusion of this additional factor.

We shape our analysis to proceed with forecasting four macroeconomic variables with forecast relevant factors. As shown in Table 1.2, selecting among all the factors included in the system results with smaller forecast losses. Comparisons of Case 1 and 4, Case 2 and 5 and Case 3 and 6 point out that variable selection leads to smaller losses. Given the small number of exceptions[10], the analysis addresses strong evidence for variable selection by the tailored information criterion

[10]The exceptions are $OLS - AL$ IP Cases 4, 5 and 6 for $\alpha = 0.1$. PI_{OLS-AL} Cases 4, 5 and 6 for $\alpha = 0.9$. PI_{AL} Case 5 for $\alpha = 0.9$

AIC$_{\mathcal{L}}$ being useful.

Our analysis is not designed to select an "optimal" α, since α is imposed by the beneficiary of the forecast i.e. the corresponding loss preferences. Yet, our results can still deliver some insight on the matter. For forecasting unemployment rate, $\alpha = 0.1$ appears to be the optimal value which leads to the smallest forecast losses for all cases. For the other three variables, $\alpha = 0.9$ results with the smallest forecast errors for all cases.

We additionally compared the $OLS-AL$ and AL forecasts with the help of the Diebold-Mariano [DM] test for predictive accuracy (Diebold and Mariano, 1995). The null hypothesis is that the expected forecast loss is equal for both procedures of interest, $\tilde{y}^{(1)}_{t+h}$ and $\tilde{y}^{(2)}_{t+h}$. The losses implied by these forecasts are $\mathcal{L}(\tilde{v}^{(1)}_{t+h})$ and $\mathcal{L}(\tilde{v}^{(2)}_{t+h})$. Under the null hypothesis, $H_0 : \mathrm{E}\left(\mathcal{L}(\tilde{v}^{(1)}_{t+h})\right) = \mathrm{E}\left(\mathcal{L}(\tilde{v}^{(2)}_{t+h})\right)$ or $H_0 : E(d_t) = 0$ where $d_t = \mathcal{L}(\tilde{v}^{(1)}_{t+h}) - \mathcal{L}(\tilde{v}^{(2)}_{t+h})$ is the loss differential, the DM test statistic is $S = \bar{d}/(\widehat{LRV}(\bar{d})/\bar{T})^{0.5} \sim N(0,1)$ where $\bar{T}$ is the number of forecast errors available for comparison and $\widehat{LRV}$ is an estimate of the asymptotic (long-run) variance of $\sqrt{\bar{T}}\bar{d}$. Since we compute differences between AL and $OLS-AL$, we may expect test statistics to be smaller than -1.645 at the 5% significance level when $OLS-AL$ is inferior.

Table 1.3. Tests of equal predictive accuracy of OLS and AL based forecasts

Alpha	Cases	DM Test PI	DM Test IP	DM Test UN	DM Test SL
0.1	Case 1	-3.07^*	-3.07^*	-2.91^*	-2.72^*
	Case 2	-3.67^*	-2.93^*	-2.85^*	-2.76^*
	Case 3	-3.46^*	-2.68^*	-2.80^*	-2.72^*
	Case 4	-2.69^*	-3.61^*	-2.83^*	-2.57^*
	Case 5	-2.70^*	-3.63^*	-2.83^*	-2.57^*
	Case 6	-2.70^*	-3.16^*	-2.83^*	-2.57^*
0.3	Case 1	-2.28^*	-1.57	-0.91	-1.50
	Case 2	-2.82^*	-1.44	-0.87	-1.85^*
	Case 3	-2.54^*	-1.22	-0.93	-1.76^*
	Case 4	-3.04^*	-2.28^*	-0.79	-1.32
	Case 5	-3.04^*	-2.28^*	-0.79	-1.32
	Case 6	-3.04^*	-2.15^*	-0.79	-1.32
0.5	Case 1	-0.29	-2.30^*	-1.62	-0.37
	Case 2	0.76	0.89	-1.85^*	1.99
	Case 3	-0.62	0.98	-1.04	0.52
	Case 4	0.92	-1.36	-0.33	0.99
	Case 5	0.92	-1.36	-0.33	0.99
	Case 6	0.92	-0.87	-0.33	1.55
0.7	Case 1	-0.48	-0.65	-1.54	-2.88^*
	Case 2	-0.27	-0.70	-1.65^*	-2.25^*
	Case 3	-0.39	-0.80	-1.50	-2.32^*
	Case 4	-0.27	-0.01	-1.92^*	-2.69^*
	Case 5	-0.27	0.03	-1.92^*	-2.45^*
	Case 6	-0.27	-0.06	-1.79^*	-2.46^*
0.9	Case 1	-2.02^*	-2.19^*	-2.93^*	-4.97^*
	Case 2	-1.73^*	-2.18^*	-3.06^*	-4.86^*
	Case 3	-1.78^*	-2.20^*	-2.97^*	-4.60^*
	Case 4	-2.51^*	-1.74^*	-3.37^*	-3.83^*
	Case 5	-2.51^*	-1.59	-3.51^*	-3.62^*
	Case 6	-2.51^*	-1.83^*	-3.13^*	-3.63^*

Notes: The null hypothesis for the test is $H_0 : E[d_t] = 0$ where $d_t = \mathcal{L}(\hat{v}_{t+h}) - \mathcal{L}(\tilde{v}_{t+h})$ with $\hat{v}_{t+h}$ the forecast errors from OLS based forecasts and $\tilde{v}_{t+h}$ the asymmetric loss forecast errors. For the one sided test with the alternative hypothesis $H_0 : E[d_t] > 0$, the test statistic should be smaller than -1.645 for 5% significance. Significant outcomes are marked with an asterisk.

Table 1.3 reports the DM statistics of the comparison between the OLS based forecasts and Asymmetric Loss based forecasts. The test statistics confirm our expectations for $\alpha = 0.1$ without any exceptions. Except for PI, test statistics are insignificant for $\alpha = 0.3$. The forecasts are not significantly better or worse for $\alpha = 0.5$ [11] but remain insignificant for $\alpha = 0.7$ except significant results for Sales. For α being 0.9 the test gives significant results for all six cases and all variables.

1.5 Concluding remarks

The forecasting literature often focusses on MSE-optimal forecasts. Yet there is evidence emphasising the relevance of more general loss functions in concrete situations. In this paper, we incorporate some aspects of forecasting under asymmetric loss functions in factor-based predictive regressions. First, we show that one may estimate predictive regressions under the relevant loss by plugging in factors extracted from a data set by means of a first-step principal components analysis. The estimated optimal forecast from the feasible regression converges in probability to the theoretical optimal forecast. Second, we address the relevance of the estimated factors by assessing whether they are forecast-relevant under *a given* loss function. To this end, we employ *tailored* information criteria and consider the factors with highest predictive powers for forecasting purposes. Moreover, we argue that principal component analysis does not always extract all relevant information: we analyze *the variability* of the predictor series and include corresponding additional information in the forecasting model, namely a factor extracted from the log-squared idiosyncratic components estimated in the first-step PCA.

We then illustrate the discussion by forecasting the Personal Income, Industrial Production, Unemployment Rate and Retail Sales series from the Stock and Watson data set. We resort to a recursive pseudo out-of-sample forecast evaluation scheme where the factors are extracted from a subset of the Stock and Watson data and used for forecasting one-year-ahead values of PI, IP, UN and SL under several asymmetric power loss functions. We compare six forecasting models (Case 1: fifteen factors; Case 2: fifteen factors and the volatility factor; Case 3: fifteen factors, the volatility factor and squared volatility factor, Case 4: selection of forecast-relevant factors among fifteen, Case 5: selection of forecast-relevant factors among sixteen, Case 6: selection of forecast-relevant factors among seventeen) for different parameter values when the $p = 2$ is fixed. Expectedly, fitting the forecasting model under the relevant loss function leads to smaller averaged losses compared to the case when we use MSE in the majority of cases. Adding

[11] The loss differential should be zero when $\alpha = 0.5$ and $p = 2$ as $OLS - AL$ and AL are identical for this particular case. The $OLS - AL$ and the AL estimators for $\alpha = 0.5$ are however computed in a different manner (via the QR decomposition for OLS and by numerical optimization for AL), hence some negligible numerical differences arise resulting in non-zero but insignificant DM statistics.

volatility information sometimes improves the forecasts. Model selection taking the relevant loss into account leads to overall best results.

Both our theoretical and empirical results underscore the importance of using forecast-relevant information by estimating factors from an auxiliary data set to exploit the additional information (i.e. the volatility factor in our case). Also relevant, if not even more so, is the issue of choosing the most relevant information for the particular loss function used to define optimality of the forecast.

Appendix 1

1.A An information criterion

Following Akaike (1973), the definition of the information criterion in form of a penalized log-likelihood leads to

$$AIC\left(k\right) = -2\ln\left(\hat{L}\left(k\right)\right) + 2k$$

with $\hat{L}\left(k\right)$ denoting the maximum of the likelihood function for model complexity k.

Suppose now that the error term in the model of interest follows an asymmetric (exponential) power distribution as characterized by Ayebo and Kozubowski (2003) and Komunjer (2007)[12] with density function

$$f\left(v\right) = \frac{\delta^{\frac{1}{\lambda}}}{\sigma\Gamma\left(1+\frac{1}{\lambda}\right)} e^{-\delta\left(\frac{1}{\alpha_*^\lambda}I(v\leq 0)+\frac{1}{(1-\alpha_*)^\lambda}I(v>0)\right)\left|\frac{v}{\sigma}\right|^\lambda}$$

where $\delta = \frac{2\alpha_*^\lambda(1-\alpha_*)^\lambda}{\alpha_*^\lambda+(1-\alpha_*)^\lambda}$. Quasi-ML estimation of a regression model assuming $v_t \sim f$ is then easily shown to be equivalent to estimation under the loss function $\mathcal{L}$ with parameters $p = \lambda$ and $\alpha = \frac{(1-\alpha_*)^p}{(1-\alpha_*)^p+\alpha_*^p}$.

After concentrating out σ, some algebra leads to

$$AIC_{\mathcal{L}}\left(k\right) = \frac{2}{p}\ln\left(\sum\mathcal{L}\left(\hat{v}_{t+h}\right)\right) + \frac{2k}{T}$$

with $\hat{v}_t$ the residuals from estimation of the predictive regression under the relevant loss $\mathcal{L}$.

This reduces to the AIC when $\mathcal{L}$ is the squared-error loss function. Note that AIC$_{\mathcal{L}}$ differs from the IC proposed by (Weiss, 1996, Section 5) in two important respects. First, Weiss focusses on comparing forecasts from models based on different loss functions, while we are interested in selecting the best forecasting model for a given loss function; second, the expression he arrives at is not scale invariant, whereas, for the loss function in (1.2), AIC$_{\mathcal{L}}$ is.

1.B Proof of Proposition 1

Note first that $\mathcal{L}$ is continuous and piecewise linear for $p = 1$, while, for $p > 1$ it is smooth with continuous and piecewise linear $p-1$st order derivative.

[12] They introduce asymmetry in the exponential power (also generalized power, or generalized error) distribution by using the method discussed in Fernandez et al. (1995). An alternative way of "skewing" the exponential power distribution is based on the approach of Azzalini (1985).

The target function is given by

$$
\begin{aligned}
&\mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) \\
&= \frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(y_{t+h} - c^* - \sum_{j=1}^{q} a_j^* y_{t-j+1} - \sum_{k=1}^{r} b_k^* \hat{f}_{t,k}\right) \\
&= \frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(v_{t+h} - (c^* - c) - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k} + \sum_{k=1}^{r} b_k^*\left(f_{t,k} - \hat{f}_{t,k}\right)\right).
\end{aligned}
$$

In a first step, we show that

$$
\begin{aligned}
&\mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) \\
&= \frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(v_{t+h} - (c^* - c) - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k}\right) + o_p(1)
\end{aligned}
$$

where the $o_p(1)$ term is uniform in t as follows.

Let

$$
q_t = v_{t+h} - (c^* - c) - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k}
$$

and $\Delta q_t = \sum_{k=1}^{r} b_k^*\left(f_{t,k} - \hat{f}_{t,k}\right)$.

For $p = 1$, $\mathcal{L}$ is Lipschitz such that

$$
\left|\mathcal{L}\left(q_t + \Delta q_t\right) - \mathcal{L}\left(q_t\right)\right| \leq C\left|\Delta q_t\right|,
$$

which can be re-written as

$$
\mathcal{L}\left(q_t + \Delta q_t\right) = \mathcal{L}\left(q_t\right) + C\xi_t
$$

where $\left|\xi_t\right| \leq \left|\Delta q_t\right|$.

For $p = 2$, use the mean value theorem, and, for $p > 2$, a Taylor expansion of order $p-1$ with the rest term in differential form, to obtain that

$$
\mathcal{L}\left(q_t + \Delta q_t\right) = \mathcal{L}\left(q_t\right) + \mathcal{L}'\left(q_t\right)\Delta q_t + \ldots + \frac{1}{(p-1)!}\mathcal{L}^{(p-1)}\left(q_t + \xi_t\right)\left(\Delta q_t\right)^{p-1}
$$

where again $|\xi_t| \leq |\Delta q_t|$. Summing up, we obtain

$$\left| \mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) - \frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(q_t + \Delta q_t\right) \right|$$
$$\leq \sum_{j=1}^{p-2} \frac{1}{j!}\frac{1}{T}\sum_{t=p+1}^{T} \left|\mathcal{L}^{(j)}\left(q_t\right)\right| \left|\left(\Delta q_t\right)^j\right| + \frac{1}{(p-1)!}\frac{1}{T}\sum_{t=p+1}^{T} \left|\mathcal{L}^{(p-1)}\left(q_t + \xi_t\right)\right| \left|\left(\Delta q_t\right)^{p-1}\right|.$$

Note that $\mathcal{L}^{(p-1)}$ is Lipschitz continuous, so we have that

$$\left|\mathcal{L}^{(p-1)}\left(q_t + \xi_t\right) - \mathcal{L}^{(p-1)}\left(q_t\right)\right| \leq C\left|\xi_t\right| \leq C\left|\Delta q_t\right|$$

and it follows that

$$\left| \mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) - \frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(q_t + \Delta q_t\right) \right|$$
$$\leq C\sum_{j=1}^{p-2} \frac{1}{T}\sum_{t=p+1}^{T} \left|\mathcal{L}^{(j)}\left(q_t\right)\right| \left|\Delta q_t\right|^j + C\frac{1}{T}\sum_{t=p+1}^{T} \left|\mathcal{L}^{(p-1)}\left(q_t\right)\right| \left|\Delta q_t\right|^{p-1} + C\frac{1}{T}\sum_{t=p+1}^{T} \left|\Delta q_t\right|^p.$$

Proposition 2 in Bai (2003) establishes that $\sup_t \left|f_{t,k} - \hat{f}_{t,k}\right| \xrightarrow{p} 0$ for all k, so we have immediately that $\sup_t \left|\Delta q_t\right|^j \xrightarrow{p} 0$, such that

$$\frac{1}{T}\sum_{t=p+1}^{T} \left|\Delta q_t\right|^p \xrightarrow{p} 0.$$

Moreover, for all $1 \leq j \leq p-1$,

$$\frac{1}{T}\sum_{t=p+1}^{T} \left|\mathcal{L}^{(p-1)}\left(q_t\right)\right| \left|\Delta q_t\right|^{p-1} \leq \sup_t \left|\Delta q_t\right|^{p-1} \frac{1}{T}\sum_{t=p+1}^{T} \left|\mathcal{L}^{(j)}\left(q_t\right)\right| \xrightarrow{p} 0$$

since $\left|\mathcal{L}^{(j)}\left(q_t\right)\right| \leq C\left|q_t\right|^j$ for suitable C, and q_t has finite pth order moments (because v_{t+h}, y_t and $f_{t,k}$ do), such that, thanks to the Markov's inequality, $\frac{1}{T}\sum_{t=p+1}^{T} \left|q_t\right|^j$ is uniformly bounded in probability.

Then, we resort to the ergodic law of large numbers to establish that

$$\frac{1}{T}\sum_{t=p+1}^{T} \mathcal{L}\left(q_t\right) \xrightarrow{p} \mathrm{E}\left(\mathcal{L}\left(q_t\right)\right)$$

pointwise in the parameter space. To this end note that y_t is a stable AR filtering of v_{t+h} and $\sum_{k=1}^{r} b_k f_{t,k}$, so $y_t, y_{t-1}, \ldots, f_{t,k}, v_{t+h}$ is a jointly stationary and ergodic process, and that the finiteness of $\mathrm{E}\left(\left|\mathcal{L}\left(q_t\right)\right|\right)$ is given since

$$\mathcal{L}\left(q_t\right) \leq C\left|q_t\right|^p,$$

where the expectation of the r.h.s. is finite whenever $\left\|q_t\right\|_p = \sqrt[p]{\mathrm{E}\left(\left|q_t\right|^p\right)}$ is finite. But Minkowski's inequality indicates that $\left\|q_t\right\|_p$ is finite whenever the L_p norm of y_t and $f_{t,k}$ is finite, which is the case given that y_t and $f_{t,k}$ have finite pth order moments.

Hence

$$\mathcal{Q}\left(a_j^*, b_k^*, c^*, a_j, b_k, c\right) \xrightarrow{p} \mathrm{E}\left(\mathcal{L}\left(v_{t+h} - \left(c^* - c\right) - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k}\right)\right)$$

pointwise. Since $\mathcal{L}$ is convex, Lemma II.1 of Andersen and Gill (1982) applies such that the above convergence is uniform on any compact set.

Finally, we only have to check that the above expectation is minimized for $a_j^* = a_j$, $b_k^* = b_k$ and $c^* = c$; given the continuity of $\mathcal{Q}$, consistency of the estimators $\tilde{a}_j$, $\tilde{b}_k$ and $\tilde{c}$ follows via the continuity of the argmin operator w.r.t. the sup norm. To this end, note that, since the generalized forecast error is a martingale difference sequence with no atom at the origin, it holds that

$$\arg\min_{v^*} \mathrm{E}\left(\mathcal{L}\left(v_{t+h} - v^* \middle| y_{t-j}, f_t\right)\right) = 0$$

uniquely, implying that, for any $v^* \neq 0$,

$$\begin{aligned} \mathrm{E}\left(\mathcal{L}\left(v_{t+h} - v^*\right)\right) &= \mathrm{E}\left(\mathrm{E}\left(\mathcal{L}\left(v_{t+h} - v^* \middle| y_{t-j}, f_t\right)\right)\right) \\ &> \mathrm{E}\left(\mathrm{E}\left(\mathcal{L}\left(v_{t+h} \middle| y_{t-j}, f_t\right)\right)\right) = \mathrm{E}\left(\mathcal{L}\left(v_{t+h}\right)\right) \end{aligned}$$

such that $\mathrm{E}\left(\mathcal{L}\left(v_{t+h} - \sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) F_{t,k}\right)\right)$ must be minimized for $\sum_{j=1}^{q}\left(a_j^* - a_j\right) y_{t-j+1} - \sum_{k=1}^{r}\left(b_k^* - b_k\right) f_{t,k} = 0$ which, with y_t and $f_{t,k}$ linearly independent stochastic processes, is only the case when $a_j^* - a_j = b_k^* - b_k = 0$ for all $1 \leq j \leq q$ and $1 \leq k \leq r$. The consistency of the forecast function follows immediately.

1.C Model selection using the LASSO

The basic flavor

The LASSO estimator minimizes here the target function

$$\sum \mathcal{L}\left(v_{t+h}^*\right) + \lambda\left\|\boldsymbol{\beta}^*\right\|_1$$

where $\boldsymbol{\beta}^*$ stacks all parameters of the predictive regression (1.10), v^*_{t+h} are the forecast errors implied by $\boldsymbol{\beta}^*$, and $\|\cdot\|_1$ stands for the L_1 vector norm. The penalty parameter λ controls the relative importance of the shrinkage term $\lambda\,\|\boldsymbol{\beta}^*\|_1$. In spite of this not being a penalized OLS regression as originally discussed by Tibshirani (1994), the essential properties of the LASSO plausibly hold: if $\mathcal{L}$ is smooth, the non-smooth nature of the L_1 constraint may still force some of the estimates to be exactly zero – which is in effect selection among the regressors; cf. Tibshirani (1994).

For the full data span, we compare now the LASSO-supported model selection with the exercise in Table 1.1. Choosing a high λ leads to a parsimonious model; the penalty parameter is decreased to allow for more flexible models. Table 1.4 gives an outline of the resulting models selected for the dependent variable Personal Income over the the full data span. We restrict ourselves to $\alpha = 0.5$ for the illustration, but otherwise this is the LASSO analog of Table 1.1. The columns of Table 1.4 show the selected factors while we consider model selection among 8 through 15 factors. In the initial stage of the step with a total of 8 factors (second column), LASSO with high penalty parameter selects the first factor. The smaller the penalty parameters gets, the more flexible the model becomes and includes other factors one at the time when wise decreasing λ. The seventh factor is chosen as the second relevant which is followed by the selection of the fifth. Eventually λ reaches 0 and LASSO chooses all 8 factors. The same is done for a total of 9 factors in the third column etc.

The main message is, again, concerning the order in which the factors are included in the forecasting model. The LASSO attaches different importance on the factors which is indeed different than the order of the PCA extraction of these factors. For instance, the second factor by PCA is the sixth most important among 8 factors as exhibited in the second column of Table 1.4. Added factors as λ decreases also show resemblance with the tailored AIC results when we contemplate more than 8 factors. Jumping right to the case with 15 factors, in the last column of the table, the steps demonstrate the relevance of the factors starting with the most parsimonious model just with the first factor. As λ gets smaller, eleventh, seventh, tenth factors and so on are chosen until the process ends with the selection of the twelfth factor, the most irrelevant one.

Table 1.4. Selection of Factors (LASSO), full data span

Stage	Added Factors							
1	1	1	1	1	1	1	1	1
2	7	7	7	11	11	11	11	11
3	5	5	10	7	7	7	7	7
4	8	8	5	10	10	10	10	10
5	3	3	8	5	5	13	13	13
6	2	2	3	8	8	5	5	5
7	6	6	2	3	3	8	8	15
8	4	9	6	2	2	3	3	8
9		4	9	6	6	2	2	3
10			4	9	9	6	14	2
11				4	4	9	6	14
12					12	4	9	6
13						12	4	9
14							12	4
15								12

Notes: The LASSO is investigated for model selection when the dependent variable is Personal Income and the loss is quadratic. See the text for details.

For the actual forecasting exercise with LASSO as model selection tool, the actual selection of the penalty parameter λ (implying the selection of a specific predictive model) is conducted with the help of the modified AIC.

Some refinements

Belloni and Chernozhukov (2013) proposed several refinements of the LASSO. The most prominent variant is the post-fit LASSO which performs as good as LASSO, in some cases even better and has smaller bias (at least for OLS-based LASSO). In addition to post-fit LASSO, Belloni and Chernozhukov proposed post-LASSO and post-threshold LASSO; the former has slightly inferior performance compared to the post-fit LASSO while the latter is beyond the scope of this exercise. Therefore, we also considered the post-fit LASSO.

Tables 1.5 and 1.6 illustrate the losses evaluated the same way as the construction of Table 1.2 (see Section 2.4) however the model selection method differs here. Basically, both LASSO and post-fit LASSO are employed instead of tailored AIC. Hence, the results of Cases 4, Case 5 and Case 6 are reported below for both methods.

For the post-fit LASSO, Tables 1.5 and 1.6 use different penalty parameters which is enabled by the different choice of a tuning parameter introduced by Belloni and Chernozhukov (2013). The penalty parameter for the former table is simulated to be high, while, for the latter, we generate the results by using a low penalty parameter which leads a more flexible model.[13] The average losses for Case 1, 2 and 3 are the same as in Table 1.2 and Case 4 LASSO, Case 5 LASSO and Case 6 LASSO results are numerically identical on the Tables 1.5 and 1.6 which are provided just for comparison. The interpretation of the results is almost similar to the case of tailored AIC.

[13]For Table 1.5, a high penalty parameter is obtained by choosing the tuning parameter low, 0.05 for this particular case. Table 1.6 is generated with a low penalty parameter which is simulated when the tuning parameter is chosen to be 0.95.

Table 1.5. Average Losses with LASSO-based model selection - Parsimonious post-fit LASSO

Alpha	Cases	PI_{OLS-AL}	PI_{AL}	IP_{OLS-AL}	IP_{AL}	UN_{OLS-AL}	UN_{AL}	SL_{OLS-AL}	SL_{AL}
0.1	Case 1	0.2536	0.2023	0.1793	0.0988	0.0092	0.0054	0.8751	0.6065
	Case 2	0.2532	0.2004	0.1845	0.1019	0.0094	0.0054	0.9153	0.6035
	Case 3	0.2520	0.2042	0.1785	0.1009	0.0096	0.0055	0.9176	0.6137
	Case 4 LASSO	0.2570	0.1861	0.1799	0.0949	0.0092	0.0052	0.8556	0.5883
	Case 5 LASSO	0.2587	0.1877	0.1847	0.0967	0.0092	0.0052	0.8535	0.5874
	Case 6 LASSO	0.2529	0.1944	0.1765	0.0966	0.0092	0.0052	0.8515	0.5934
	Case 4 post-fit LASSO	0.2386	0.1791	0.1690	0.0851	0.0085	0.0051	0.7947	0.5443
	Case 5 post-fit LASSO	0.2385	0.1791	0.1676	0.0846	0.0085	0.0051	0.7943	0.5440
	Case 6 post-fit LASSO	0.2389	0.1823	0.1697	0.0898	0.0085	0.0051	0.7946	0.5444
0.3	Case 1	0.2369	0.2232	0.1695	0.1494	0.0098	0.0092	0.8668	0.8149
	Case 2	0.2365	0.2218	0.1728	0.1532	0.0100	0.0092	0.8859	0.8119
	Case 3	0.2366	0.2231	0.1695	0.1520	0.0102	0.0093	0.8945	0.8247
	Case 4 LASSO	0.2425	0.2241	0.1716	0.1477	0.0098	0.0090	0.8271	0.7843
	Case 5 LASSO	0.2391	0.2188	0.1719	0.1469	0.0099	0.0091	0.8160	0.7716
	Case 6 LASSO	0.2374	0.2177	0.1694	0.1482	0.0098	0.0090	0.8176	0.7760
	Case 4 post-fit LASSO	0.2234	0.2079	0.1580	0.1335	0.0091	0.0086	0.7859	0.7466
	Case 5 post-fit LASSO	0.2234	0.2079	0.1560	0.1316	0.0091	0.0086	0.7859	0.7466
	Case 6 post-fit LASSO	0.2235	0.2083	0.1576	0.1350	0.0091	0.0086	0.7859	0.7466
0.5	Case 1	0.2201	0.2201	0.1598	0.1598	0.0105	0.0105	0.8586	0.8586
	Case 2	0.2198	0.2198	0.1611	0.1611	0.0106	0.0106	0.8564	0.8564
	Case 3	0.2212	0.2212	0.1604	0.1604	0.0108	0.0108	0.8714	0.8714
	Case 4 LASSO	0.2233	0.2233	0.1599	0.1599	0.0103	0.0103	0.8093	0.8093
	Case 5 LASSO	0.2229	0.2229	0.1593	0.1593	0.0105	0.0105	0.8079	0.8079
	Case 6 LASSO	0.2229	0.2229	0.1591	0.1591	0.0104	0.0104	0.8086	0.8086
	Case 4 post-fit LASSO	0.2104	0.2104	0.1400	0.1400	0.0098	0.0098	0.7792	0.7792
	Case 5 post-fit LASSO	0.2104	0.2104	0.1399	0.1399	0.0098	0.0098	0.7792	0.7792
	Case 6 post-fit LASSO	0.2111	0.2111	0.1409	0.1409	0.0098	0.0098	0.7792	0.7792
0.7	Case 1	0.2034	0.2006	0.1500	0.1419	0.0111	0.0098	0.8503	0.7890
	Case 2	0.2031	0.2016	0.1494	0.1403	0.0111	0.0099	0.8269	0.7889
	Case 3	0.2058	0.2035	0.1514	0.1404	0.0114	0.0102	0.8484	0.8047
	Case 4 LASSO	0.2030	0.2011	0.1506	0.1439	0.0110	0.0099	0.8044	0.7413
	Case 5 LASSO	0.2036	0.2015	0.1490	0.1440	0.0112	0.0100	0.8010	0.7392
	Case 6 LASSO	0.2052	0.2029	0.1500	0.1428	0.0112	0.0101	0.8037	0.7411
	Case 4 post-fit LASSO	0.1975	0.1950	0.1255	0.1257	0.0104	0.0090	0.7726	0.7099
	Case 5 post-fit LASSO	0.1975	0.1950	0.1254	0.1253	0.0105	0.0091	0.7726	0.7099
	Case 6 post-fit LASSO	0.1987	0.1958	0.1321	0.1294	0.0105	0.0091	0.7726	0.7099
0.9	Case 1	0.1866	0.1537	0.1403	0.0848	0.0118	0.0064	0.8420	0.5554
	Case 2	0.1864	0.1555	0.1377	0.0834	0.0117	0.0066	0.7975	0.5728
	Case 3	0.1903	0.1577	0.1423	0.0839	0.0119	0.0066	0.8253	0.5922
	Case 4 LASSO	0.1836	0.1500	0.1411	0.0877	0.0116	0.0064	0.8062	0.5314
	Case 5 LASSO	0.1868	0.1511	0.1386	0.0890	0.0119	0.0066	0.8001	0.5321
	Case 6 LASSO	0.1872	0.1516	0.1415	0.0891	0.0119	0.0065	0.8016	0.5383
	Case 4 post-fit LASSO	0.1845	0.1515	0.1186	0.0817	0.0111	0.0055	0.7659	0.4984
	Case 5 post-fit LASSO	0.1845	0.1515	0.1166	0.0810	0.0112	0.0056	0.7659	0.4984
	Case 6 post-fit LASSO	0.1862	0.1526	0.1216	0.0810	0.0112	0.0056	0.7659	0.4984

Notes: The construction of this table follows that of Table 1.2. The only difference is that there are two selection methods reported for Case 3 and 4. First method uses LASSO by Tibshirani (1994) with AIC$_{\mathcal{L}}$ choice of the penalty parameter λ, while the second uses post-fit LASSO by Belloni and Chernozhukov (2013). Post-fit LASSO require pre specification of a tuning parameter which is used to simulate the penalty parameter. Here, the penalty is chosen to be high through the tuning parameter (0.05) which leads to a parsimonious model.

Table 1.6. Average Losses with LASSO-based model selection - Flexible post-fit LASSO

Alpha	Cases	PI_{OLS-AL}	PI_{AL}	IP_{OLS-AL}	IP_{AL}	UN_{OLS-AL}	UN_{AL}	SL_{OLS-AL}	SL_{AL}
	Case 1	0.2536	0.2023	0.1793	0.0988	0.0092	0.0054	0.8751	0.6065
	Case 2	0.2532	0.2004	0.1845	0.1019	0.0094	0.0054	0.9153	0.6035
	Case 3	0.2520	0.2042	0.1785	0.1009	0.0096	0.0055	0.9176	0.6137
	Case 4 LASSO	0.2570	0.1861	0.1799	0.0949	0.0092	0.0052	0.8556	0.5883
0.1	Case 5 LASSO	0.2587	0.1877	0.1847	0.0967	0.0092	0.0052	0.8535	0.5874
	Case 6 LASSO	0.2529	0.1944	0.1765	0.0966	0.0092	0.0052	0.8515	0.5934
	Case 4 post-fit LASSO	0.2529	0.1907	0.1737	0.0925	0.0088	0.0052	0.7998	0.5732
	Case 5 post-fit LASSO	0.2515	0.1904	0.1722	0.0888	0.0087	0.0052	0.7946	0.5683
	Case 6 post-fit LASSO	0.2511	0.1889	0.1696	0.0924	0.0088	0.0052	0.7948	0.5762
	Case 1	0.2369	0.2232	0.1695	0.1494	0.0098	0.0092	0.8668	0.8149
	Case 2	0.2365	0.2218	0.1728	0.1532	0.0100	0.0092	0.8859	0.8119
	Case 3	0.2366	0.2231	0.1695	0.1520	0.0102	0.0093	0.8945	0.8247
	Case 4 LASSO	0.2425	0.2241	0.1716	0.1477	0.0098	0.0090	0.8271	0.7843
0.3	Case 5 LASSO	0.2391	0.2188	0.1719	0.1469	0.0099	0.0091	0.8160	0.7716
	Case 6 LASSO	0.2374	0.2177	0.1694	0.1482	0.0098	0.0090	0.8176	0.7760
	Case 4 post-fit LASSO	0.2351	0.2178	0.1664	0.1460	0.0096	0.0090	0.8165	0.7943
	Case 5 post-fit LASSO	0.2352	0.2182	0.1676	0.1458	0.0096	0.0090	0.8202	0.7983
	Case 6 post-fit LASSO	0.2374	0.2205	0.1647	0.1456	0.0097	0.0090	0.8303	0.8065
	Case 1	0.2201	0.2201	0.1598	0.1598	0.0105	0.0105	0.8586	0.8586
	Case 2	0.2198	0.2198	0.1611	0.1611	0.0106	0.0106	0.8564	0.8564
	Case 3	0.2212	0.2212	0.1604	0.1604	0.0108	0.0108	0.8714	0.8714
	Case 4 LASSO	0.2233	0.2233	0.1599	0.1599	0.0103	0.0103	0.8093	0.8093
0.5	Case 5 LASSO	0.2229	0.2229	0.1593	0.1593	0.0105	0.0105	0.8079	0.8079
	Case 6 LASSO	0.2229	0.2229	0.1591	0.1591	0.0104	0.0104	0.8086	0.8086
	Case 4 post-fit LASSO	0.2163	0.2163	0.1624	0.1624	0.0103	0.0103	0.8264	0.8264
	Case 5 post-fit LASSO	0.2172	0.2172	0.1601	0.1601	0.0103	0.0103	0.8263	0.8263
	Case 6 post-fit LASSO	0.2193	0.2193	0.1605	0.1605	0.0103	0.0103	0.8211	0.8211
	Case 1	0.2034	0.2006	0.1500	0.1419	0.0111	0.0098	0.8503	0.7890
	Case 2	0.2031	0.2016	0.1494	0.1403	0.0111	0.0099	0.8269	0.7889
	Case 3	0.2058	0.2035	0.1514	0.1404	0.0114	0.0102	0.8484	0.8047
	Case 4 LASSO	0.2030	0.2011	0.1506	0.1439	0.0110	0.0099	0.8044	0.7413
0.7	Case 5 LASSO	0.2036	0.2015	0.1490	0.1440	0.0112	0.0100	0.8010	0.7392
	Case 6 LASSO	0.2052	0.2029	0.1500	0.1428	0.0112	0.0101	0.8037	0.7411
	Case 4 post-fit LASSO	0.2007	0.1995	0.1483	0.1409	0.0109	0.0096	0.8350	0.7524
	Case 5 post-fit LASSO	0.2012	0.1995	0.1498	0.1435	0.0110	0.0097	0.8322	0.7506
	Case 6 post-fit LASSO	0.2038	0.2022	0.1510	0.1434	0.0110	0.0097	0.8497	0.7690
	Case 1	0.1866	0.1537	0.1403	0.0848	0.0118	0.0064	0.8420	0.5554
	Case 2	0.1864	0.1555	0.1377	0.0834	0.0117	0.0066	0.7975	0.5728
	Case 3	0.1903	0.1577	0.1423	0.0839	0.0119	0.0066	0.8253	0.5922
	Case 4 LASSO	0.1836	0.1500	0.1411	0.0877	0.0116	0.0064	0.8062	0.5314
0.9	Case 5 LASSO	0.1868	0.1511	0.1386	0.0890	0.0119	0.0066	0.8001	0.5321
	Case 6 LASSO	0.1872	0.1516	0.1415	0.0891	0.0119	0.0065	0.8016	0.5383
	Case 4 post-fit LASSO	0.1842	0.1520	0.1338	0.0843	0.0116	0.0063	0.8472	0.5338
	Case 5 post-fit LASSO	0.1853	0.1523	0.1302	0.0840	0.0118	0.0064	0.8452	0.5303
	Case 6 post-fit LASSO	0.1870	0.1557	0.1439	0.0896	0.0118	0.0065	0.8547	0.5408

Notes: Here, the penalty is chosen to be low through the tuning parameter (0.95) which leads to a more flexible model. See Table1.5 for further details.

1.D Data description

All series are from the Global Insights Basic Economics Database, unless the source is listed (in parentheses) as TCB (The Conference Boards Indicators Database) or AC (authors calculation based on Global Insights or TCB data). Transformation codes are indicated in the column named *Tcode*. 1 indicates levels of the series, 2 and 3 denote the first and the second differences, respectively and 4 means logarithm, 5 and 6 respectively indicate the first and the second differences of the logarithm of the series.

	Mnemonic	Short Desc	Tcode	Description
1	ypr	PI	5	Personal Income (Ar, Bil. Chain 2000 $)
2	a0m051	PI less transfers	5	Personal Income Less Transfer Payments (Ar, Bil. Chain 2000 $)
3	cons_r	Consumption	5	Real Consumption (Ac) A0M224/Gmdc
4	mtq	M&T sales	5	Manufacturing And Trade Sales (Mil. Chain 1996 $)
5	a0m059	Retail sales	5	Sales Of Retail Stores (Mil. Chain 2000 $)
6	ips10	IP: total	5	Industrial Production Index - Total Index
7	ips11	IP: products	5	Industrial Production Index - Products, Total
8	ips299	IP: final prod	5	Industrial Production Index - Final Products
9	ips12	IP: cons gds	5	Industrial Production Index - Consumer Goods
10	ips13	IP: cons dble	5	Industrial Production Index - Durable Consumer Goods
11	ips18	iIP:cons nondble	5	Industrial Production Index - Nondurable Consumer Goods
12	ips25	IP:bus eqpt	5	Industrial Production Index - Business Equipment
13	ips32	IP: matls	5	Industrial Production Index - Materials
14	ips34	IP: dble mats	5	Industrial Production Index - Durable Goods Materials
15	ips38	IP:nondble mats	5	Industrial Production Index - Nondurable Goods Materials
16	ips43	IP: mfg	5	Industrial Production Index - Manufacturing (Sic)
17	ips307	IP: res util	5	Industrial Production Index - Residential Utilities
18	ips306	IP: fuels	5	Industrial Production Index - Fuels
19	pmp	NAPM prodn	1	NAPM Production Index (Percent)
20	utl11	Cap util	2	Capacity Utilization (Mfg)
21	lhel	Help wanted indx	2	Index Of Help-Wanted Advertising In Newspapers (1967=100;Sa)
22	lhelx	Help wanted/emp	2	Employment: Ratio; Help-Wanted Ads:No. Unemployed Clf
23	lhem	Emp CPS total	5	Civilian Labor Force: Employed, Total (Thous.,Sa)
24	lhnag	Emp CPS nonag	5	Civilian Labor Force: Employed, Nonagric.Industries (Thous.,Sa)
25	lhur	U: all	2	Unemployment Rate: All Workers, 16 Years & Over (%,Sa)
26	lhu680	U: mean duration	2	Unemploy.By Duration: Average(Mean)Duration In Weeks (Sa)
27	lhu5	U $<$ 5 wks	5	Unemploy.By Duration: Persons Unempl.Less Than 5 Wks (Thous.,Sa)
28	lhu14	U 5-14 wks	5	Unemploy.By Duration: Persons Unempl.5 To 14 Wks (Thous.,Sa)
29	lhu15	U 15+ wks	5	Unemploy.By Duration: Persons Unempl.15 Wks + (Thous.,Sa)
30	lhu26	U 15-26 wks	5	Unemploy.By Duration: Persons Unempl.15 To 26 Wks (Thous.,Sa)

	Mnemonic	Short Desciption	Tcode	Description
31	lhu27	U 27+ wks	5	Unemploy.By Duration: Persons Unempl.27 Wks + (Thous,Sa)
32	claimuii	UI claims	5	Average Weekly Initial Claims, Unemploy. Insurance (Thous.)
33	ces002	Emp: total	5	Employees On Nonfarm Payrolls - Total Private
34	ces003	Emp: gds prod	5	Employees On Nonfarm Payrolls - Goods-Producing
35	ces006	Emp: mining	5	Employees On Nonfarm Payrolls - Mining
36	ces011	Emp: const	5	Employees On Nonfarm Payrolls - Construction
37	ces015	Emp: mfg	5	Employees On Nonfarm Payrolls - Manufacturing
38	ces017	Emp: dble gds	5	Employees On Nonfarm Payrolls - Durable Goods
39	ces033	Emp: nondbles	5	Employees On Nonfarm Payrolls - Nondurable Goods
40	ces046	Emp: services	5	Employees On Nonfarm Payrolls - Service-Providing
41	ces048	Emp: TTU	5	Employees On Nonfarm Payrolls - Trade, Transportation, And Utilities
42	ces049	Emp: wholesale	5	Employees On Nonfarm Payrolls - Wholesale Trade
43	ces053	Emp: retail	5	Employees On Nonfarm Payrolls - Retail Trade
44	ces088	Emp: fire	5	Employees On Nonfarm Payrolls - Financial Activities
45	ces140	Emp: Govt	5	Employees On Nonfarm Payrolls - Government
46	ces151	Avg hrs	1	Avg Wkly Hours, Prod Wrkrs, Nonfarm - Goods-Producing
47	ces155	Overtime: mfg	2	Average Weekly Hours Of Production Or Nonsupervisory Workers On Private Nonfar
48	a0m001	Avg hrs: mfg	1	Average Weekly Hours, Mfg. (Hours)
49	pmemp	NAPM empl	1	NAPM Employment Index (Percent)
50	hsfr	HStarts: Total	4	Housing Starts:Nonfarm(1947-58);Total Farm&Nonfarm(1959-) (Thous.,Sa)
51	hsne	HStarts: NE	4	Housing Starts:Northeast (Thous.U,Sa).
52	hsmw	HStarts: MW	4	Housing Starts:Midwest (Thous.U,Sa).
53	hssou	HStarts: South	4	Housing Starts:South (Thous.U,Sa)
54	hswst	HStarts: West	4	Housing Starts:West (Thous.U,Sa)
55	hsbr	BP: total	4	Housing Authorized: Total New Priv Housing Units (Thous.,Saar)
56	hsbne	BP: NE	4	Houses Authorized By Build. Permits:Northeast (Thous.U,Sa)
57	hsbmw	BP: MW	4	Houses Authorized By Build. Permits:Midwest (Thous.U,Sa)
58	hsbsou	BP: South	4	Houses Authorized By Build. Permits:South (Thous.U,Sa)
59	hsbwst	BP: West	4	Houses Authorized By Build. Permits:West (Thous.U,Sa)
60	pmi	PMI	1	Purchasing Managers Index (Sa)
61	pmno	NAPM new ordrs	1	NAPM New Orders Index (Percent)
62	pmdel	NAPM vendor del	1	NAPM Vendor Deliveries Index (Percent)
63	pmnv	NAPM Invent	1	NAPM Inventories Index (Percent)
64	a1m008	Orders: cons gds	5	Mfrs New Orders, Consumer Goods And Materials (Bil. Chain 1982 $)
65	a0m007	Orders: dble gds	5	Mfrs New Orders, Durable Goods Industries (Bil. Chain 2000 $)

	Mnemonic	Short Desciption	Tcode	Description
66	a0m027	Orders: cap gds	5	Mfrs New Orders, Nondefense Capital Goods (Mil. Chain 1982 $)
67	a1m092	Unf orders: dble	5	Mfrs Unfilled Orders, Durable Goods Indus. (Bil. Chain 2000 $)
68	a0m070	M&T invent	5	Manufacturing And Trade Inventories (Bil. Chain 2000 $)
69	a0m077	M&T invent/sales	2	Ratio, Mfg. And Trade Inventories To Sales (Based On Chain 2000 $)
70	fm1	M1	6	Money Stock: M1(Curr,Trav.Cks,Dem Dep,Other Ck Able Dep)(Bil$,Sa)
71	fm2	M2	6	Money Stock:M2(M1+O Nite Rps,Euro$,G/P&B/D Mmmfs&Sav&Sm Time Dep(Bil$,
72	fmscu	M3	6	Money Stock: M3(M2+Lg Time Dep,Term Rp S&Inst Only Mmmfs)(Bil$,Sa)
73	fm2_r	M2 (real)	5	Money Supply - M2 In 1996 Dollars (Bci)
74	fmfba	MB	6	Monetary Base, Adj For Reserve Requirement Changes(Mil$,Sa)
75	fmrra	Reserves tot	6	Depository Inst Reserves:Total,Adj For Reserve Req Chgs(Mil$,Sa)
76	fmrnba	Reserves nonbor	6	Depository Inst Reserves:Nonborrowed,Adj Res Req Chgs(Mil$,Sa)
77	fclnbw	C&I loans	6	Commercial & Industrial Loans Oustanding In 1996 Dollars (Bci)
78	fclbmc	C&I loans	1	Wkly Rp Lg Com L Banks:Net Change Com L & Indus Loans(Bil$,Saar)
79	ccinrv	Cons credit	6	Consumer Credit Outstanding - Nonrevolving(G19)
80	ccipy	Inst cred/PI	2	Ratio, Consumer Instalment Credit To Personal Income (Pct.)
81	fspcom	S&P 500	5	S&P S Common Stock Price Index: Composite (1941-43=10)
82	fspin	S&P: indust	5	S&P S Common Stock Price Index: Industrials (1941-43=10)
83	fsdxp	S&P div yield	2	S&P S Composite Common Stock: Dividend Yield (% Per Annum)
84	fspxe	S&P PE ratio	5	S&P S Composite Common Stock: Price-Earnings Ratio (%,Nsa)
85	fyff	FedFunds	2	Interest Rate: Federal Funds (Effective) (% Per Annum,Nsa)
86	cp90	Commpaper	2	Commercial Paper Rate (Ac)
87	fygm3	3 mo T-bill	2	Interest Rate: U.S.Treasury Bills,Sec Mkt,3-Mo.(% Per Ann,Nsa)
88	fygm6	6 mo T-bill	2	Interest Rate: U.S.Treasury Bills,Sec Mkt,6-Mo.(% Per Ann,Nsa)
89	fygt1	1 yr T-bond	2	Interest Rate: U.S.Treasury Const Maturities,1-Yr.(% Per Ann,Nsa)
90	fygt5	5 yr T-bond	2	Interest Rate: U.S.Treasury Const Maturities,5-Yr.(% Per Ann,Nsa)
91	fygt10	10 yr T-bond	2	Interest Rate: U.S.Treasury Const Maturities,10-Yr.(% Per Ann,Nsa)
92	fyaaac	Aaabond	2	Bond Yield: Moody S Aaa Corporate (% Per Annum)
93	fybaac	Baa bond	2	Bond Yield: Moody S Baa Corporate (% Per Annum)
94	scp90	CP-FF spread	1	Cp90-Fyff
95	sfygm3	3 mo-FF spread	1	Fygm3-Fyff
96	sfygm6	6 mo-FF spread	1	Fygm6-Fyff
97	sfygt1	1 yr-FF spread	1	Fygt1-Fyff
98	sfygt5	5 yr-FF spread	1	Fygt5-Fyff
99	sfygt10	10 yr-FF spread	1	Fygt10-Fyff
100	sfyaaac	Aaa-FF spread	1	Fyaaac-Fyff

	Mnemonic	Short Desciption	Tcode	Description
101	sfybaac	Baa-FF spread	1	Fybaac-Fyff
102	exrus	Ex rate: avg	5	United States;Effective Exchange Rate(Merm)(Index No.)
103	exrsw	Ex rate: Switz	5	Foreign Exchange Rate: Switzerland (Swiss Franc Per U.S.$)
104	exrjan	Ex rate: Japan	5	Foreign Exchange Rate: Japan (Yen Per U.S.$)
105	exruk	Ex rate: UK	5	Foreign Exchange Rate: United Kingdom (Cents Per Pound)
106	exrcan	EX rate: Canada	5	Foreign Exchange Rate: Canada (Canadian $ Per U.S.$)
107	pwfsa	PPI: fin gds	6	Producer Price Index: Finished Goods (82=100,Sa)
108	pwfcsa	PPI: cons gds	6	Producer Price Index:Finished Consumer Goods (82=100,Sa)
109	pwimsa	PPI: int matls	6	Producer Price Index:Intermed Mat.Supplies & Components(82=100,Sa)
110	pwcmsa	PPI: crude matls	6	Producer Price Index:Crude Materials (82=100,Sa)
111	psccom	Commod: spot price	6	Spot Market Price Index:Bls & Crb: All Commodities(1967=100)
112	pw102	Sens matls price	6	Index Of Sensitive Materials Prices (1990=100)(Bci-99A)
113	pmcp	NAPM com price	1	NAPM Commodity Prices Index (Percent)
114	punew	CPI-U: all	6	Cpi-U: All Items (82-84=100,Sa)
115	pu83	CPI-U: apparel	6	Cpi-U: Apparel & Upkeep (82-84=100,Sa)
116	pu84	CPI-U: transp	6	Cpi-U: Transportation (82-84=100,Sa)
117	pu85	CPI-U: medical	6	Cpi-U: Medical Care (82-84=100,Sa)
118	puc	CPI-U: comm.	6	Cpi-U: Commodities (82-84=100,Sa)
119	pucd	CPI-U: dbles	6	Cpi-U: Durables (82-84=100,Sa)
120	pus	CPI-U: services	6	Cpi-U: Services (82-84=100,Sa)
121	puxf	CPI-U: ex food	6	Cpi-U: All Items Less Food (82-84=100,Sa)
122	puxhs	CPI-U: ex shelter	6	Cpi-U: All Items Less Shelter (82-84=100,Sa)
123	puxm	CPI-U: ex med	6	Cpi-U: All Items Less Medical Care (82-84=100,Sa)
124	gmdc	PCE defl	6	Pce,Impl Pr Defl:Pce (1987=100)
125	gmdcd	PCE defl: dlbes	6	Pce,Impl Pr Defl:Pce; Durables (1987=100)
126	gmdcn	PCE defl: nondble	6	Pce,Impl Pr Defl:Pce; Nondurables (1996=100)
127	gmdcs	PCE defl: services	6	Pce,Impl Pr Defl:Pce; Services (1987=100)
128	ces275	AHE: goods	6	Average Hourly Earnings Of Production Or Nonsupervisory Workers On Private No
129	ces277	AHE: const	6	Average Hourly Earnings Of Production Or Nonsupervisory Workers On Private No
130	ces278	AHE: mfg	6	Average Hourly Earnings Of Production Or Nonsupervisory Workers On Private No
131	hhsntn	Consumer expect	2	U. Of Mich. Index Of Consumer Expectations(Bcd-83)

CHAPTER 2

Interpreting Latent Dynamic Factors by Threshold FAVAR Model

2.1 Introduction

Data availability has evolved rapidly in the recent years. Hundreds of variables are readily available for use. However, using such large data sets introduce a challenge by bringing model specification and estimation problems along. Employing factor models can deal with these seemingly adverse issues. Factor models beneficially adapt large information sets to the analysis by providing a convenient tool to reduce dimensions and to extract information. True specifications of the models that researchers are interested in have been successfully accomplished thanks to factor models, particularly by including the large information sets available to policy makers. However, interpreting factors is still a black box. To this purpose, we propose a factor-augmented VAR model by introducing a latent threshold which induces the factor loadings onto zero when the factors are found irrelevant given the estimated threshold level. The shut down rate of the factor loadings, which we can construct by observing the frequency of factor loadings induced onto zero, reveals the relationship between factors and macroeconomic variables.

Researchers might simply want to use large information sets to make use of all the relevant information available. To overcome the difficulty of using many indicators up to some extend, vector autoregressions (VARs) are designed to include more than one evolving variable, as a generalization of autoregression models. VARs have been acknowledged as successfully identifying the direction and the magnitude of monetary shocks since the time they were proposed by Bernanke and Blinder (1992) and Sims (1992b).

Despite VARs' common use, especially among macroeconomists, the relatively small number of macroeconomic variables in VARs cannot capture all the necessary information and might cause omitted variable bias. Another point worth noting in VARs is the selection of the variables. There are generally different measures of the same series, e.g. output, inflation or unemployment. Even for the same country these series can differ but all might include some information that others do not. Unfortunately, VAR results heavily depend on the choice of these series. Furthermore, adding more variables to VARs creates degrees of freedom issues. In this matter, factor models play an important role in enabling us to use large information sets by extracting common factors. These factors are latent variables capturing the common fluctuations in the data. One can imagine the set of factors as the summary of the information in that particular data set. Therefore, the curse of dimensionality does not occur in factor-augmented models.

Nevertheless, factor models alone cannot explain the effects of, e.g. monetary policy, shocks on all macroeconomic variables. However, due to the nature of factor models, macroeconomic shocks cannot be traced back to the variables. Therefore, Bernanke et al. (2005) combined factor models with VARs to be able to use both large information sets and explain the effects of monetary shocks on various indicators. This new model, factor augmented VAR (FAVAR) can be used to assess vast data sets and to observe impulse response functions of all variables.

Factor models, and consequently FAVAR models, are useful at a cost. It is unfortunately not possible to interpret the factors which actually might have been beneficial to link them to macroeconomic indicators. Belviso and Milani (2006) acknowledged this problem and proposed the Structural FAVAR (SFAVAR) model. Their SFAVAR model divides the large information set into subgroups of particular economic activities. Only one factor is extracted from each category. Thereby this factor is simply associated with the corresponding group. Certainly others have attempted to interpret factors by using different approaches, e.g. Negro and Otrok (2008), Ludvigson and Ng (2009b,c), Bork (2009).

Nakajima and West (2013b) proposes the threshold procedure on the factor loadings. The adapted approach aims to use threshold structure for modeling in dynamic factor volatility models as an extension in Bayesian sparsity modeling. In this paper, we extend their model and propose a latent threshold FAVAR model. The adaptation is based on the following: the factors to be extracted from the data may not be relevant for some time periods. Here, some of the loadings are induced to zero for the particular time periods unless they are above a threshold level which is endogenously estimated. This strategy implicitly allows us to detect the factor loadings that are frequently or rarely shut down for specific macroeconomic variables.

Overall, we ask the following questions: What if a factor loading is shut down particularly for one or more groups of macroeconomic variables throughout time and only a few (preferably one) of the factors are related to particular variables? Can we infer which factor(s) might be related to one particular subgroup of data? The unique intent of this paper is to estimate latent threshold FAVAR model to develop a new method to assign economic interpretation to estimated factors. A likely alternative to SFAVAR, this analogous model comes with the difference of a latent threshold structure on the factor loadings matrix. The main objective is to detect the irrelevancy of some factors for certain time periods, especially for some variables.

The data driven shrinkage clearly defines a more sparse model. Therefore, this allows us to identify the factors which might carry information about some subgroups of the data or the factors which are totally irrelevant for some. We do so by inquiring the frequency of the shut down and surviving factor loadings to infer the relationships between the factors and the variables, and relate them. The strategy we use clarifies the interpretation of the factors by approaching these questions from a different angle compared to Belviso and Milani (2006)'s SFAVAR approach. Our approach does *not* require a prespecification of the data set. We boldly aim to let whole data decide on which variables factors have effects. Therefore, the approach we propose here is more general in the sense of detecting the factors related to certain subgroups in the data set.

The proposed method may seem similar to the time varying parameter FAVAR (TVP–FAVAR) where the factor loadings and some other parameters are allowed to differ over time as in Korobilis (2009), Liu et al. (2011), Baumeister et al. (2010) and Eickmeier et al. (2011a,b) among numerous others. In the time varying parameter models, the point when the loadings become sufficiently small and, hence, irrelevant is not easily identifiable since we do not have a strict measure of the threshold under which the factors become redundant. The factor loadings in this paper are also time varying. However our approach concentrates more on a specific time varying loadings scheme to interpret the factors. The threshold structure enables us to observe this measure and induce the loadings to zero for irrelevant factors on associated time periods.

We estimate the model with Bayesian techniques where we use a data set constructed by quarterly macroeconomic indicators running from 1964:Q1 to 2013:Q1. The first set of our results presents the survival rates which we observe the frequency of shut downs in factor loadings. The factors are mainly assigned to one group of macroeconomic indicators such as unemployment, inflation/finance or real economy. The second set of findings depicts the impulse response functions. The responses of factors to monetary contraction are generally of expected sign. Impulse response functions of factors against shocks on factors and of individual variables against interest rate shock generally are in line with economic theory suggests.

The paper proceeds as follows. Section 2.2 introduces our model and summarizes the Bayesian estimation along with the restrictions we impose. The data set we use is discussed in section

2.3. Section 2.4 presents the results for number of the factors to be used and elaborates on interpretation of the factors. The details of impulse response functions are displayed in section 2.5. Section 2.6 concludes and presents the future work. All other relevant information, including the impulse response functions which are not discussed throughout the main sections, different identification restrictions and the data description are given in Appendix.

2.2 The Model

2.2.1 Model Specification

The model used in this paper comprises a VAR system along with a factor model. Let X_t be a $N \times 1$ vector of observed macroeconomic series. These series form an information set in factor analysis. We seek to observe the impact of the observable policy variable, $m \times 1$ vector Y_t, on the large data set of economic activity, X_t. Hence, monetary economists frequently take Y_t as Federal Funds Rate (FFR), as in this paper, but in practice this is not a restriction. We can also have several (policy) variables in Y_t. The unobserved variables are factors f_t, $k \times 1$ vector, and the time varying factor loading matrix Λ_t of dimension $N \times k$.

The model has 3 main equations: a state equation where f_t and Y_t follow a VAR(q) process, a measurement equation which illustrates how the large data set X_t is related to the latent factors f_t and the policy variables Y_t, and lastly the autoregressive process for the latent threshold factor loadings. Typical FAVAR model has first two parts. The threshold part is borrowed from Nakajima and West (2013a,b).

Assume the joint process of the factors and the policy variable can be represented in the state equation as a reduced VAR,

$$\begin{bmatrix} f_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} f_t \\ Y_t \end{bmatrix} + \varepsilon_t, \quad \text{for } t = 1, \ldots, T, \tag{2.1}$$

where $\varepsilon_t \sim \mathcal{N}(0, \Sigma)$ and $\Phi(L) = \Phi_1 L + \Phi_2 L^2 + \cdots + \Phi_q L^q$ is a lag polynomial of order q with each Φ_j is $K \times K$ matrix for $j = 1, \ldots, q$ satisfying stationarity, where $K = k + m$. We need to solve the structural VAR form to obtain impulse-response functions which will be discussed in the following sections.

The state equation cannot be estimated by itself since the factors are unobservable. A small number of factors, $k << N$, is extracted from the data as the representatives of the common fluctuations and used in the state equation to interact with Y_t. Therefore we need the following measurement equation,

$$X_t = c_t + \Lambda_t f_t + \gamma Y_t + e_t, \quad \text{for } t = 1, \ldots, T, \tag{2.2}$$

where e_t is $N \times 1$ vector of idiosyncratic components such that $e_t \sim \mathcal{N}(0, \Omega_t)$ where Ω_t is $N \times N$ diagonal time varying covariance matrix and $E(e_t \,|\, F_t, Y_t) = 0$ with $E(e_{jt}, e_{lt}) = 0$ for all $j, l = 1, \ldots, N$ and $j \neq l$. We assume that the diagonal elements of matrix Ω_t follow a stochastic volatility process, that is, $\Omega_t = \text{diag}\{e^{h_{1,t}}, \ldots, e^{h_{N,t}}\}$ is in the form of

$$h_t = \mu_h + \alpha_h(h_{t-1} - \mu_h) + v_{ht}$$

with $v_{ht} \sim \mathcal{N}(0, V_h)$ where both α_h and V_h are $N \times N$ diagonal matrices and $h_t = (h_{1,t}, \ldots, h_{N,t})'$. The time varying intercept follows a stationary autoregressive process

$$c_t = \mu_c + \alpha_c(c_{t-1} - \mu_c) + v_{ct}$$

with $v_{ct} \sim \mathcal{N}(0, V_c)$ where both α_c and V_c are $N \times N$ diagonal matrices. The time varying constant and variance help us capturing the changes in the data over time, especially when the time varying parameters tend to create unstable results, e.g. as in the Great Recession period.

Factors are representatives of the variations in the data however their relevance might depend on the particular time periods and therefore change over time. Hence, the factor loadings in our model are *not* left unrestricted but instead represented by a threshold structure. Intuitively, the idea is to examine the relative importance of the factors in each time period. This specific representation enables us to observe whether factor loadings are below a threshold and which should be induced to zero for the associated time periods.

To exploit the above insight, we stack all the non-zero elements in the loadings matrix Λ_t.[1] Let us denote each non-zero element of Λ_t as λ_{jt}. Then the threshold structure on the factor loadings is,

$$\lambda_{jt} = \beta_{jt} \mathbb{1}(|\beta_{jt}| \geq \delta_j), \quad \text{for } j = 1, \ldots, p,$$

where $p = (N - k + 1)k$ is the number of the non-zero loadings, $\mathbb{1}(\cdot)$ denotes the indicator function, $\delta_j \geq 0$ is the latent threshold for $j = 1, \ldots, p$ which is to be estimated. The latent time varying parameter vector $\beta_t = (\beta_{1t}, \ldots, \beta_{pt})$ follows stationary VAR(1) model

$$\beta_t = \mu_\beta + \alpha_\beta(\beta_{t-1} - \mu_\beta) + v_{\beta t}, \tag{2.3}$$

where $v_{\beta t} \sim \mathcal{N}(0, V_\beta)$, μ_β is $p \times 1$, α_β and V_β are both $p \times p$ diagonal matrices. The AR coefficient of β_{jt} satisfies the stationarity of AR(1) processes for each factor loading, i.e. $|\alpha_{\beta j}| < 1$. Suffice

[1] The zero elements are due to the identification restrictions, which will be explained in the next sections.

it to say, we assume that the errors of different equations are jointly normal and independent. That is, $(e_t, \varepsilon_t, v_{\beta t}, v_{ct}, v_{ht})' \sim \mathcal{N}(0, \text{diag}(\Omega_t, \Sigma, V_\beta, V_c, V_h))$, where $\text{diag}(\cdot)$ creates a block diagonal matrix. Moreover, all of the covariance matrices except Σ are diagonal. The Appendix provides details on the priors and the posteriors of the parameters.

This threshold factor model has some advantages over continuous time-varying loading models and Markov switching (MS) loading models. In continuous time-varying loadings framework, the (time-varying) importance of a factor can be inferred through the magnitude of the loading over time. However, there is no scale which indicates how small λ_{jt} should be so that the factor is considered redundant. Hence, when a factor becomes important is very subjective. In the threshold model, on the other hand, the threshold is estimated. Therefore the data decide when a factor should be included in the analysis. In an MS setup, one can have two (or a finite number of) regimes for the loadings: significant and insignificant regimes. Both MS and the threshold model behave similarly when a loading is shut-down to 0. However, for the time periods when a factor is significant, the threshold model allows continuous loadings which ensures a better fit than MS loading models.

2.2.2 Bayesian Estimation

The estimation of the parameters and latent processes of the factor model relies mostly on the results of Nakajima and West (2013b). We employ the Markov chain Monte Carlo (MCMC) method to estimate the joint distribution of the unobserved variables. The full posterior density conditional on the data is $p(\Psi_{0:T}, \delta, \theta, \gamma, \Phi, \Sigma | X_{(1:N,1:T)}, Y_{(1:m,1:T)})$ where $\Psi_{0:T} = \{c_{0:T}, \beta_{0:T}, f_{1:T}, h_{0:T}\}$ are the latent time-varying processes, $\delta = \{\delta_1, \ldots, \delta_p\}$ are the latent thresholds for each non-zero element of the loading matrix, $\theta = \{\theta_c, \theta_h, \theta_\beta\}$ where $\theta_g = \{\mu_g, \alpha_g, \sigma_g^2\}$ for $g \in \{c, h, \beta\}$, γ is $N \times m$ matrix of measurement equation parameter, Φ and Σ are the VAR parameters, and $\{X_{(1:N,1:T)}, Y_{(1:m,1:T)}\}$ is the data X_{it} and Y_{jt} for $i = 1, \ldots, N$, $j = 1, \ldots, m$ and $t = 1, \ldots, T$.

The estimation of $c_{0:T}$ and $f_{1:T}$ can be performed by forward filtering backward sampling algorithm conditional on the hyperparameters, the time-varying volatility and the data. In this paper we use Carter and Kohn (1994) algorithm which draws the time series of the latent process in a state space representation. The volatility process $h_{0:T}$ is sampled by standard MCMC techniques developed for univariate stochastic volatility models conditional on the measurement equation parameters and the data. The parameters θ_c and θ_h are sampled easily after conditioning on $c_{0:T}$ and $h_{0:T}$, respectively, as in simple univariate AR(1) models.

We use Metropolis–Hasting algorithm to draw $\delta, \beta_{0:T}, \theta_\beta$. The estimation of these parameters is deeply analyzed in Nakajima and West (2013a). The candidate for β is drawn from a distribution as if there is no threshold. The draws for θ_β are required to be compatible with the threshold parameters because the prior and the posterior of δ depends on θ_β. We performed 25000 iterations and discarded the first 20000 draws as burn-in period. Convergence of most of the parameters is achieved. Some details are given in the Appendix, but for further details readers should refer Nakajima and West (2013a,b).

2.2.3 Identification Restrictions for Factors

As widely covered in the literature, the estimation of the true factors cannot be achieved. Instead only the space spanned by the factors can be estimated. Moreover, unless we apply some restrictions, we cannot identify the factors and loadings separately. In other words, for any given factor f and loadings Λ the following observational equivalence holds: $\Lambda f = \Lambda R R^{-1} f = \tilde{\Lambda}\tilde{f}$ for invertible $k \times k$ matrix R, i.e., same results can be achieved by two different sets of factors and factor loadings. Thus we need to fix the rotation of the factors, namely fixing the matrix R, by putting k^2 restrictions.

In Principal Component Analysis, a statistical method to extract factors from data sets, the most common restrictions are to assume ff'/T being identity matrix ($k(k+1)/2$ restrictions) and $\Lambda\Lambda'$ being diagonal ($k(k-1)/2$ restrictions). However different restrictions have been adopted by both dynamic factor and FAVAR models. For instance Bernanke et al. (2005) and numerous others following their work restrict the top $k \times k$ block of Λ to be identity. Some of the dynamic factor model papers such as Aguilar and West (2000) and Nakajima and West (2013b) restrict the top $k \times k$ block of Λ to be lower triangular with unit diagonals which leads $k(k+1)/2$ restrictions. Additionally they restrict the covariance matrix of the factors, Σ to be diagonal which brings along $k(k-1)/2$ more restrictions.

We believe that restricting the covariance matrix of the factors by forcing for unit diagonals and no correlation between factors is a very strong restriction. The impulse response functions are generated through the covariance matrix. Thus, such restrictions are indeed undesirable. Furthermore, we would like to keep the factor loadings as free as possible since the interpretation of the factors are based on the loadings. In our paper, we imposed diagonality on the lower $k \times k$ block of Λ and set the diagonals of the top $k \times k$ block of Σ to be one. Restricting the bottom part of the factor loadings has some intuitive grounds. The ordering of our data set allows us to assume that each of the last k variables is only explained by one factor.[2] Moreover, setting the variances of the unobserved factors, the corresponding diagonal elements of Σ, as 1 is just a normalization. Leaving off–diagonal elements of the covariance matrix of the factors unrestricted indicates that correlation among factors is allowed, e.g. the correlation between so called ‘inflation factor’ and ‘interest rate factor’ is left unrestricted in our analysis. The

[2]The corresponding variables in the data set are the credit variables.

restrictions on both covariance matrix, k, and the factor loadings, $k^2 - k$, provide us the total number of restrictions, k^2, we need for identification.

2.3 The Data

Factor models entail large information sets. Our data consist of 158 US macroeconomic aggregates and are inspired by Stock and Watson (2005) (SW) data set. The original SW data set and its modified versions have been used by numerous papers, such as Belviso and Milani (2006) and Ludvigson and Ng (2009b,c). In the latter, the authors touch upon the interpretation of factors and 131 monthly series in their data cover the time span of January 1964 - December 2007. Nevertheless, their data set is governed most by couple of specific groups of macroeconomic aggregates. For instance, the real activity variables are plenty whereas the financial market variable are just handful. We are fond of a more even distribution of the variables over the subgroups. Therefore we update and extended the SW data set. Although the original SW data set is monthly, we prefer to work with a quarterly data set to keep the computation tractable. Hence the resulting data set is from 1964:Q1 to 2013:Q1.[3] The number of lags in the VAR(q) is taken to be 4 throughout the analysis. Yet, the model yields similar results under different choice of lags.

We do not require any ex ante categorization of the data. However, we can benefit from looking at it in detail and also reporting the results in accordance with the different classes of variables. The subgroups and the corresponding number of variables are shown in Table 2.1 below. The data set for factor extraction includes 157 variables. The last variable, Federal Funds Rate is used as the policy variable thus it is not included in the data set from which we extract the factors.

The analysis requires all series to be stationary. This is ensured by taking differences or logarithms of the series and in some cases both. Adding more series into the data and the longer time span now require different transformations than SW's. The resulting codes are presented in the data description.

[3] Appendix presents the full data description including the data source. Most of the series are taken from St. Louis Fed Economic Research database (henceforth FRED) unless otherwise indicated.

Table 2.1. Subgroups in the Data Set

Macroeconomic Subgroups	Number of Variables
Production	20
(Un)Employment	27
Housing	13
Interest Rate	15
Inflation	29
Finance	13
Money	22
Expectations	7
Credit	11
Federal Funds Rate (FFR)	1

Notes: Appendix explains which series form these categories.

2.4 Results

We employ a Bayesian framework to extract the factors and estimate the hyperparameters. To do so first requires to determine the exact number of factors in the data. The next step is to analyze the factor loadings over time to assign an economic meaning to the factors .

2.4.1 Number of Factors in the Data

All factor related models require an initial step of determining the number of factors. There are statistical ways to seek the optimal number. Among all, the most frequently used is the information criteria for static factors proposed by Bai and Ng (2002). The crucial point in determining the optimal number is to realize that different time spans might offer different number of factors. Table 2.2 shows the results of a naive inspection on this matter.

The table presents how many factors are suggested by the information criteria for the corresponding time span of the data set. The data until the end of 2000 suggest 6 factors. However adding just the first quarter of 2001 into the time span changes the suggested number of factors. This dramatic change is not because of a sudden appearance of an actual meaningful factor. Instead, probably, there is nonlinearity caused by immediate changes in the data set, such as

Table 2.2. Number of Factors for Different Time Spans

Time Range	Number of Factors
1964Q1 - 2000Q4	6
1964Q1 - 2001Q1	7
1964Q1 - 2007Q4	7
1964Q1 - 2008Q1	8
1964Q1 - 2013Q1	8

the dot–com bubble in the beginning of 2001 for this particular case. The same notion can be observed in the third and fourth rows. Even though the information criteria suggests 7 factors until the end of 2007, adding the first quarter of 2008 leads to another factor, in this case due to the Great Recession. As said, the additional factors given by information criteria after extreme data movements do possibly capture the nonlinearity. Hence caution should be taken before treating these factors as latent variables although they survive the information criteria.

The Bai and Ng (2002) information criteria suggest that there are 8 factors in our data for the whole time span. The FAVAR model of Stock and Watson (2005) used 7 factors, only some of which were later shown to accurately construct the forecast error decomposition for individual series. Analogously, Ludvigson and Ng (2009b,c) used SW data set and extracted 8 factors as suggested by the information criteria. We, similar to Stock and Watson (2005), use 7 factors in this paper. The results of the subsequent sections show that only 5 to 6 factors are assigned economic meanings.[4] This also supports the fact that the immediate appearance of the additional factors is *artificial*. The remaining 'unmeaningful' factors are generally shut down.

2.4.2 Interpreting the Factors

Given that the factor loadings are shut down for so-called irrelevant time periods, we can observe the remaining (non-zero) loadings. This enables us to relate the factors and variables to particular data groups. If a factor's loadings are rarely induced to zero only for a specific group of macro variables, we will link that factor to the corresponding data group. The interpretation of factors depends on the 'survival rate' of the process β_{jt}. The survival rate aims to show how frequently the factor loadings are above the estimated threshold, i.e., *not* shut down to zero, and therefore the corresponding factors are relevant. We take this ratio by averaging both over simulations and time periods. Mathematically, survival rate of the j^{th} loading is $1/(TS)\sum_{t,s}\mathbb{1}(\lambda_{jt}^{(s)} \neq 0)$ where S is the number of simulations after burn-in period and $\lambda_{jt}^{(s)}$ is the s^{th} iteration of MCMC estimate of λ_{jt}. This is one of the ways of interpreting factors, which we pursue in this paper. Another would be to obtain a time-varying survival rate by averaging only over the simulations

[4] The same analysis was also repeated for 8 factors but there were no considerable changes in the results. Similarly, only 5 to 6 factors are found meaningful.

and checking the time series of loadings but interpreting the factors would be comparably harder in this case.

We introduce the subgroups of the data in Section 4 even though we treat the data as a whole for the MCMC. We ultimately intend to attach the factors to these different subgroups. Table 2.3 demonstrates the survival rates of all 7 factors for each of these subgroups. The rows indicate the average the survival rates of the top 60% of the factor loadings for the corresponding factors. This is just an adaptation for the ease of interpretability. The selection of the top percentile does *not* change the results but makes the interpretation straightforward.

The bold numbers emphasize the highest survival rates of the corresponding factors. For the production variables, for instance, the first factor is *not* shut down 63% of the time. Over time and simulations, this signifies that the first factor is above the estimated threshold with 63% probability. The fourth factor has by far the highest survival rate, 78%, among others for production. One factor might be related to other categories of the data as well, e.g. the fourth factor is also influential on housing variables with 85% survival rate. Production and housing are two highly related economic indicators hence the fourth factor can be processed as the real activity factor and is now called as 'Real' as an abbreviation.

Table 2.3. Survival Rates of the Factor Loadings

	$\mathbf{f_1}$	$\mathbf{f_2}$	$\mathbf{f_3}$	$\mathbf{f_4}$	$\mathbf{f_5}$	$\mathbf{f_6}$	$\mathbf{f_7}$
	Emp	*InfFn*	—	*Real*	*Expc*	—	*IntR*
Production	0.63	0.32	0.40	**0.78**	0.36	0.56	0.58
(Un)employment	**0.85**	0.21	0.16	0.42	0.64	0.09	0.37
Housing	0.47	0.17	0.13	**0.85**	0.10	0.11	0.55
Interest Rate	0.12	0.04	0.09	0.44	0.18	0.06	**0.51**
Inflation	0.33	**0.67**	0.31	0.25	0.27	0.36	0.52
Finance	0.38	**0.57**	0.09	0.33	0.12	0.27	0.43
Money	0.22	0.22	0.21	0.29	0.18	**0.37**	0.36
Expectations	0.59	0.22	0.05	0.29	**0.75**	0.25	0.38

Following the above mentioned analogy, we mark the first factor as employment factor, 'Emp'. In our framework, one should be careful about what a factor is truly capturing. The (un)employment partition of the data includes variables for both unemployment and employment. Can we know for sure whether the employment factor is really an employment factor or rather an unemployment factor? Visual inspection helps us to determine the actual interpretation of this factor.[5] We can simply check the correlations of every single variable with the employment factor.

[5] To identify the nature of this factor we can also put some sign restrictions on factor loadings at the beginning of the analysis.

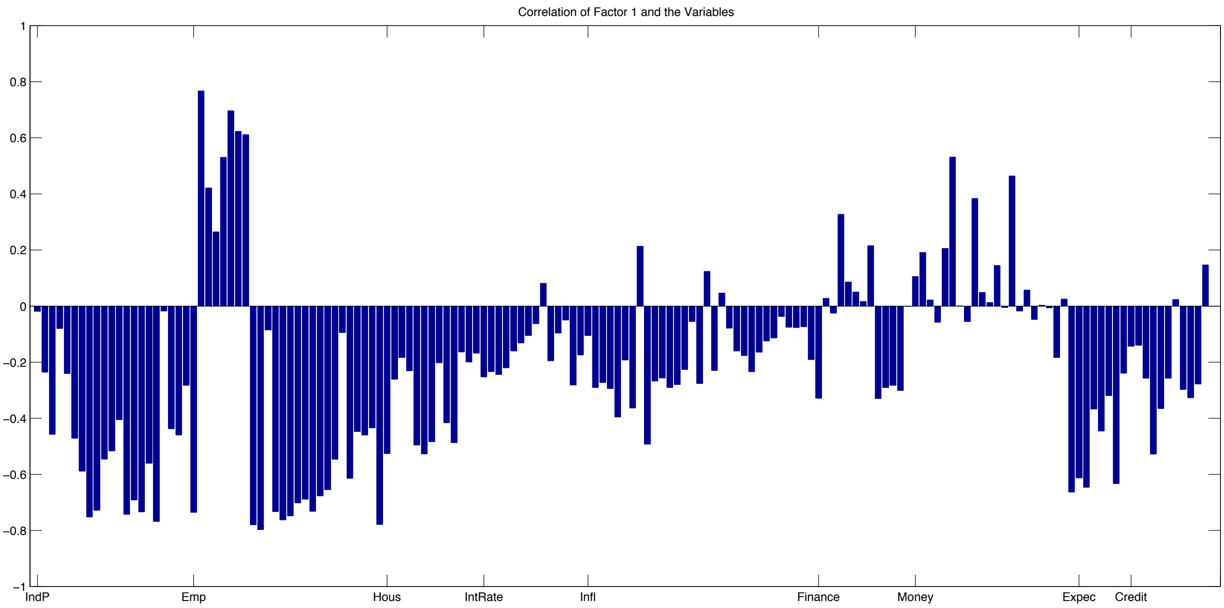


Figure 2.1. The correlation between the variables and the first factor

The positive correlations accumulated in Figure 2.1 correspond to the unemployment variables. Other variables in this same data category exhibit negative relationships with the first factor. Moreover, most of the variables (such as production, housing, expectation) are negatively correlated with this factor. Therefore, this factor can safely be identified as the unemployment factor.[6]

The second factor loads on inflation and financial variables. We cannot distinctly name this factor due to the difficulty of differentiating the effects of inflation and financial variables, hence it is indicated as 'InfFn'. The third factor is the most insignificant factor among all. This also supports the idea that some factors might be generated artificially due to capturing the nonlinearity in the data. Hence, this factor does not carry any essential information and can be left without a specific name.

The fifth factor clearly explains the expectation variables hence is indicated as 'Expc'. Expectation measures are highly related to other subgroups in the data. Stock and Watson (2005) included these indexes into the corresponding subgroups. For instance the ISM Production Index in our expectation data group is included in the real activity variables in SW data set. Nevertheless, we are able to find a strongly distinctive factor associated with the expectation variables. The existence of this factor should not be ignored in our case.

Money related variables have not been assigned to a particular factor with confidence. Even

[6]When we observe the impulse response functions of the factors after an unemployment shock in the following sections, this notion also becomes more clear.

though the most significant factor for these variables is factor 6, it might not be a conclusive result thereby it brings this factor into question. The last factor very distinctively loads on interest rate and real economy variables. It is not surprising that one factor affects more than one group as in the case of the second factor. Yet, we call the last factor as the interest rate factor.

The restrictions imposed to the model fix the rotation of the factors, i.e. we are choosing basis functions for the space spanned by the factors. Papers which forcefully assign meaning to the factors (by extracting one factor from a subgroup for instance) might end up having factors more than the dimension of the true factor space. Therefore, we believe that some of these extracted factors are either orthogonal to the true factor space or a linear combination of the true factors. According to our results here and those of similar papers', there are only 5 to 6 factors in this data set. Whichever different identification schemes we use for the estimation, we could not find any factor that explains credit variables. That is, there is no 'Credit' factor in the data based on our results from several different restrictions.[7]

2.5 Impulse Response Functions

This section presents the impulse response functions of the factors and some selected variables to particular shocks. The graphs here and in the appendix are evaluated by identifying the system with Cholesky decomposition. Attached meanings on the factors would enable us to impose better VAR identification restrictions. Cholesky decomposition is represented here just for computational advantages. The confidence bands of the impulse response functions is 68% instead of a 95% confidence interval. We prefer this due to the sampling uncertainty coming along with the estimation of the factors. Appendix provides the impulse response functions of other factors.

Figure 2.2 reports the responses of the factors to a 1 unit shock on FFR. The last of the 8 plots in each figure presents response of FFR itself. The responses are generally of the expected sign. The contractionary monetary policy shock has a relatively positive impact on unemployment factor, consistent with what economic theory suggests. Inflation and financial market factor might seem insignificant according to the response scales but has a small downward tendency at the beginning. Immediate response of the financial variables might cause this behavior.

The FFR shock has almost no effect on the third factor as the figure suggests. This supports the appearance of artificial factors as discussed in the previous sections. The real economy factor

[7] Appendix provides details on the results when we impose different identification restrictions.

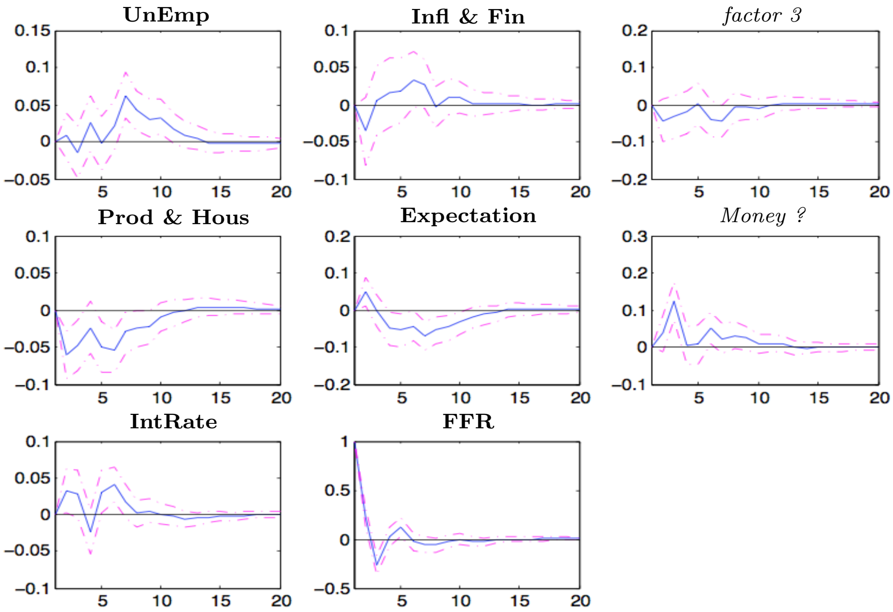


Figure 2.2. The responses of the factors and FFR to a 1 unit shock on FFR.

(we can also think of it as production and housing factor) declines. Namely, positive shock to FFR leads a drop in GDP. Expectations factor has a small upward adjustment first but then its response becomes negative, consistent with the deteriorating expectations following monetary contraction. The recovery period for expectations factor is almost the same as the real activity factor. The money factor responds positively and stays that way until the effect slowly fades. The corresponding series in the data include reserve aggregates. Therefore observing an increase in the money factor is intuitive. Lastly, interest rate factor has an upward tendency in general which is a natural response after a monetary contraction.

It is worthwhile to discuss the responses of the factors to the shocks on other factors. This is one of the crucial conveniences of FAVAR models. Interpretable factors might help to make sense of some dynamics in accordance with the impulse response functions. As an example, we concentrate on the impulse response functions of the factors and the FFR when there is an one unit positive shock to unemployment factor. The resulting responses are displayed in Figure 2.3. A sudden jump in unemployment decreases inflation and finance factor over time. In addition, impulse responses support our expectations of observing a drop in the real activity and expectations factors. Moreover, the money factor and FFR are also negatively affected by this shock whereas the response of interest rate has an upward move in the first quarters.

Another advantage of FAVAR models is that we can observe the impulse response functions of the individual variables. This provides a more intensive check on the model specification. Hence

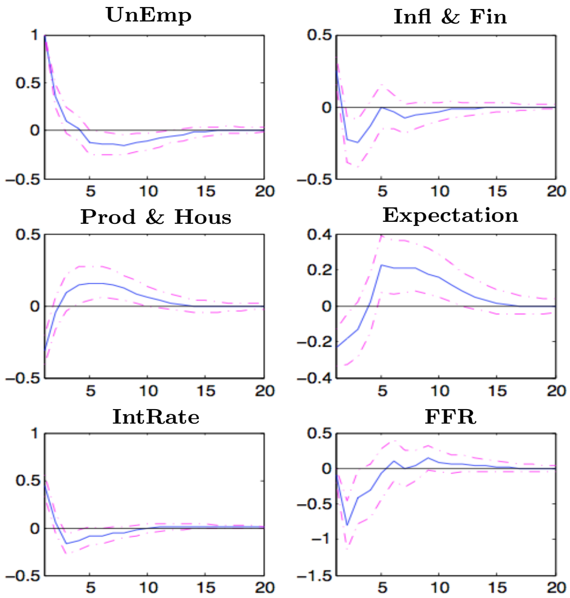


Figure 2.3. The responses of the factors and FFR to a 1 unit shock on unemployment factor.

we analyze the responses of various macroeconomic measures against a one unit contractionary monetary shock. We have a selection of different types of variables chosen from the subgroups of the data. The ordering of these variables on the data set are given next to the variable names on Figure 2.4.

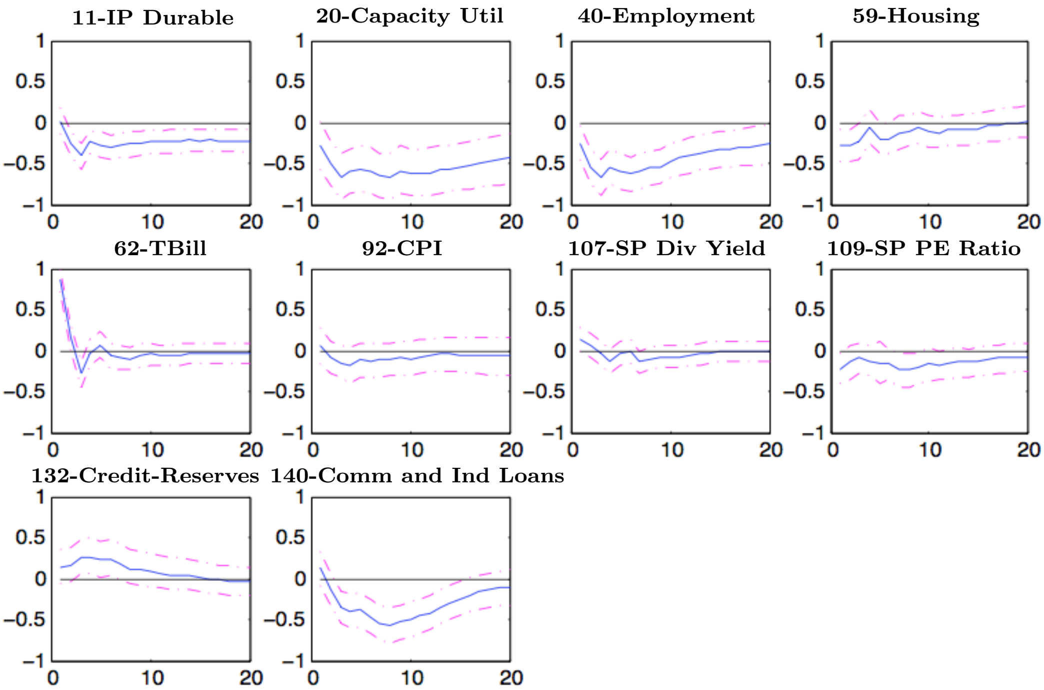


Figure 2.4. The responses of the variables to a 1 unit shock on *FFR*.

The responses of the variables are as anticipated: contradictory monetary shock leads a decline in industrial production, capacity utilization for manufacturing replies with a drop, the employment measure in the third plot has a downward adjustment along with the decrease in housing starts in the fourth plot, the FFR shock increases the 3-month Treasury Bill interests which is very similar to FFR response, similar to the findings of Bernanke et al. (2005). Dividend yields first exhibit an upward move however drop over time. Price earnings ratio of S&P declines, credit variables are expected to have a downward adjustment after a positive monetary shock.

In theory, monetary tightening should decrease the prices. However, as first identified by Sims (1992b), VAR literature suffers from a phenomenon so called *price puzzle* where prices commonly respond with an increase. That is artificially created by impulse response functions of VAR models and does not reflect what theory suggests. One of the novelties of FAVAR models is to eliminate price puzzle by making use of large data sets. In our model, CPI reacts slightly positively at the first quarter but the response becomes negative afterwards. Therefore we can infer that this model eliminates the price puzzle while this response might seem insignificant.

The results which are not displayed here are available upon request. Surely, responses of some variables are inconsistent with what the theory suggests, both in magnitude and sign. However, numerous variables are exhibiting the foreseen responses in all data categories. Moreover, the factors that we easily manage to interpret and name, such as unemployment, expectations or real activity, react as expected.

2.6 Concluding Remarks

The recent literature has focused on the techniques to efficiently use large information sets. Combining vector autoregressions with factor models is a relatively recent but very fruitful method in this regard. However, factor augmented VAR models are not designed to interpret the extracted factors. In this paper, we attempt to designate an economic meaning to the factors through a latent threshold FAVAR model.

We apply a Bayesian approach to extract the factors, interpret them according to their factor loadings, and employ a VAR analysis to observe impulse response functions of the various measures. For the identification of the factors, we need to impose some restrictions on both loadings and covariance matrices of the factors. These would be altered according to the question at hand. We intend to keep the data as a whole and employ the proposed FAVAR model. As hard as this could be for the factor interpretation, the proposed threshold structure serve our purposes well in this respect. Observing the factor loadings assists to identify irrelevant factors through the estimated threshold. Moreover, some factors seem to be *more* associated with some data categories.

The empirical evidence suggests that most factors could be related to certain subcategories of the data. Although Bai and Ng (2002) information criteria suggests the use of 8 factors for our data set, we are able to find 5 to 6 meaningful factors, e.g. real activity factor, unemployment factor. Unmeaningful factors might indicate the nonlinearity in the data which occurs after extreme economic activities, such as crises.

There are couple areas that might benefit from this approach. The potential implementation of the model, among many others, is two fold. First, it can be used on the stress testing front by performing structural analysis. Recently, central banks heavily invest on their stress testing framework alongside stress test scenarios published every year. Federal Reserve, for instance, published its 2015 severe adverse scenario where the unemployment increases by 4 percentage points, real GDP is 4.5% lower than its level in the third quarter of 2014 and CPI reaches 4.3%, see Board of Governers of the Federal Reserve System (2015). Bank of England, see Bank of England (2014), published a tail risk scenario starts with an initial shock to productivity which leads to the monetary policy response where the Bank Rate rising about 4%. Following these, unemployment rate rises to 12% and 35% fall in house prices is observed and eventually real GDP growth troughs at about 3.5%. Calibrating these numbers is only the one side of the coin. The other is the need to investigate where shocks originate. From the macroeconomic perspective, the effects of two shocks that come from different sources should have different impacts on the big scale of the economy. For instance, a real GDP fall originating from financial sector shock should have different impacts on the economy, both qualitatively and quantitatively, and different transmission mechanism than a *same size* fall in real GDP driven by a shock arising from unemployment or housing market. Calibrating the variations in macroeconomic indicators under stress should account for where shocks arise from even if they eventually materialize a same size change. Our approach can identify initial shocks by using interpretable factors which carry information on specific sectors of the economy and help gauging the ultimate numbers to be used in stress scenarios.

Second, this method can be easily extended to perform small open economy analysis. The first possible implication of this extension is to exploit the effects of the monetary contraction/expansion in a large open economy to a small open economy. Especially recently, this channel attracts more attention due to the uncertainty that might arise in small open economies, such as Canada, United Kingdom, as a response to a change in the US interest rate. With the proposed method, we have an official tool to investigate the transmission of the monetary policy from one country to another by also capturing the features of different sectors in each country. Similarly, we can explore the interconnectedness of two countries' financial sectors and/or housing sectors etc. We can easily study the propagation mechanism of, for example, a financial shock to the US economy on other countries along with the magnitude, duration and persistence of this particular shock.

The paper is open to some extensions. We seek to obtain the results under different restrictions, such as different structural VAR restrictions. They might lead to better impulse responses. Looking for *the best* factor identification restrictions might yield *the most* meaningful factors. Forecasting of particular macroeconomic series can be performed by using the proposed model. A noteworthy extension is to repeat this exercise with different data sets although compiling such data sets might be overwhelming. More micro-oriented series, such as consumption-saving measures and various indexes, or different geographical variables would create very large data sets and these can be analyzed with the aid of this model.

Appendix 2

2.A Priors and Posteriors

Prior and posterior specifications and MCMC mostly rely on Nakajima and West (2013a). This section is designed to analyze all in detail however readers can refer the original source if needed.

2.A.1 Priors

For $g \in \{\beta, c, h\}$ the priors of the parameters are as follows

$$\begin{aligned}
\mu_{i,g} &\sim \mathcal{N}(\mu_{i0}, \omega_{i0}^2) \\
(\alpha_{i,g}+1)/2 &\sim Beta(\alpha_{01}, \alpha_{02}) \\
\sigma_{i,g}^{-2} &\sim \mathcal{G}(v_{0i}/2, V_{0i}/2) \\
\beta_{i1}|\theta_\beta &\sim \mathcal{N}(\mu_{i,\beta}, \sigma_{i,\beta}^2/(1-\alpha_{i,\beta}^2)) \\
d_i|\theta_\beta &\sim \mathcal{U}\left(0, |\mu_{i,\beta}| + K\nu_i\right),
\end{aligned}$$

where $\nu_i^2 = \sigma_{i,\beta}^2/(1-\alpha_{i,\beta}^2)$ and $\sigma_{i,\beta}^2$ is the i^{th} diagonal element of V_β. Basically, the term ν_i^2 is the unconditional variance of β_{it}.

2.A.2 MCMC Estimation Steps

To perform MCMC, we use Gibbs sampling, and Metropolis-Hasting (MH) algorithm for variables related to the threshold δ. Here is the outline and some details of the MCMC estimation.

Sampling β:

The process $\beta_{0:T}$ is sampled by Metropolis-within-Gibbs sampling method. In particular, MH sampling is used for β_t conditional on β_{-t} and $\{\theta_\beta, \delta, h_{1:T}, f_{1:T}, Y_{1:T}, X_{1:T}\}$ for $t = 1, \ldots, T$. If there was no threshold, we could have easily sampled β_t's by using Kalman filter type algorithm. Hence, in the accept-reject algorithm, β_t^* which is sampled from a hypothetically no-threshold model is used as a proposal. Note that Ω_t has 0 in the off-diagonals, thus the variables in each row of the measurement equation is uncorrelated over i. That is, we can sample each row of Λ_t

independently from other rows. The conditional posterior of $k \times 1$ vector β_t under this case is $\mathcal{N}(\beta_t|m_t, M_t)$ where $i = 1, \ldots, N$ and for $t = 2 : T-1$

$$\begin{aligned} M_t^{-1} &= e^{-h_{it}} f_t f_t' + V_\beta^{-1}(I + \alpha_\beta' \alpha_\beta) \\ m_t &= M_t[e^{-h_{it}} f_t \tilde{X}_{it} + V_\beta^{-1}\{\alpha_\beta(\beta_{t-1} - \beta_{t+1}) + (I - 2\alpha_\beta + \alpha_\beta'\alpha_\beta)\mu_\beta\}] \end{aligned}$$

for $t = 1$ and $t = T$

$$\begin{aligned} M_1^{-1} &= e^{-h_{i1}} f_1 f_1' + V_{\beta,0}^{-1} + V_\beta^{-1}(I + \alpha_\beta'\alpha_\beta) \\ m_1 &= M_1[e^{-h_{i1}} f_1 \tilde{X}_{i1} + V_{\beta,0}^{-1}\mu_\beta + V_\beta^{-1}\alpha_\beta\{\beta_2 - (I - \alpha_\beta)\mu_\beta\}] \end{aligned}$$

$$\begin{aligned} M_T^{-1} &= e^{-h_{iT}} f_T f_T' + V_\beta^{-1} \\ m_T &= M_T[e^{-h_{iT}} f_T \tilde{X}_{iT} + V_\beta^{-1}\{\alpha_\beta\beta_{T-1} - (I - \alpha_\beta)\mu_\beta\}], \end{aligned}$$

where $V_{\beta,0}$ is the unconditional variance of β_t and $\tilde{X}_{it} = X_{it} - \gamma_i' Y_t$

The acceptance probability is

$$\alpha(\beta_t, \beta_t^*) = \min\left\{1, \frac{\mathcal{N}(\tilde{X}_{it}|f_t'\lambda_t^*, \exp(h_{it}))\mathcal{N}(\beta_t|m_t, M_t)}{\mathcal{N}(\tilde{X}_{it}|f_t'\lambda_t, \exp(h_{it}))\mathcal{N}(\beta_t^*|m_t, M_t)}\right\}.$$

Sampling δ:

The posterior distribution of δ_i is conditioned on $(k-1)\times 1$ vector δ_{-i} and $\{\theta_\beta, h_{1:T}, f_{1:T}, Y_{1:T}, X_{1:T}\}$. The threshold is also sampled by MH algorithm. The proposal is drawn from the conditional prior distribution $\delta_i^* \sim U\left(|\mu_i| + K\nu_i\right)$. The acceptance probability is

$$\alpha(\delta_i, \delta_i^*) = min\left\{1, \prod_{t=1}^{T} \frac{\mathcal{N}(\tilde{X}_{it}|f_t'\lambda_t^*, \exp(h_{it}))}{\mathcal{N}(\tilde{X}_{it}|f_t'\lambda_t, \exp(h_{it}))}\right\}.$$

The parameter K is a tuning parameter. It determines how large the threshold can be, thus in return, it determines the shut-down frequency of β. Nakajima and West (2013a) suggested $K = 3$ based on simulation performances, that is the threshold is drawn from a 3-standard-deviation interval. Our estimation results were pretty robust to changes in K - we estimated the model with $K \in \{1.65, 2, 3\}$.

Sampling $\{\mu_\beta, \alpha_\beta, \sigma_{i,\beta}^{-2}\}$:

These are the parameters associated with the autoregressive process for β_t. The posteriors of these parameters are typical except that they are truncated on a set where the parameter draws are compatible with the upper bound of the threshold: $D_i = \{\delta_i < |\mu_{i\beta}| + K\nu_i\}$.

The posterior density of $\mu_{i\beta}$ is $p(\mu_{i\beta}|\alpha_{i\beta}, \sigma^2_{i\beta}, \beta_{i,1:T}, \delta_i) \propto \mathcal{TN}_{D_i}(\mu_{i\beta}|\hat{\mu}_i, \hat{\omega}^2_i)(|\mu_{i\beta}| + K\nu_i)^{-1}$ where $\mathcal{TN}_{D_i}$ denotes the density of truncated normal on the set D_i, and

$$\hat{\omega}^2_i = \left\{\frac{1}{\omega^2_{i0}} + \frac{(1-\alpha^2_i) + (T-1)(1-\alpha_1)^2}{\sigma^2_{iV}}\right\}^{-1}$$

$$\hat{\mu}_i = \hat{\omega}^2_i \left\{\frac{\mu_{i0}}{\omega^2_{i0}} + \frac{(1-\alpha^2_i)\beta_{i1} + (1-\alpha_i)\sum_{t=1}^{T-1}(\beta_{i,t+1} - \alpha_i\beta_{it})}{\sigma^2_{iV}}\right\}.$$

Acceptance rate for the candidate which is drawn from the conditional posterior density is $\min\left\{1, \frac{|\mu_{i\beta}|+K\nu_i}{|\mu^*_{i\beta}|+K\nu_i}\right\}$.

The conditional posterior density of $\alpha_{i\beta}$ is

$$p(\alpha_{i\beta}|\mu_{i\beta}, \sigma^2_{i\beta}, \beta_{i,1:T}, \delta_i) \propto Beta(\alpha_{i\beta})(1-\alpha^2_{i\beta})^{1/2}\mathcal{TN}_{(-1,1)\times D_i}(\hat{\alpha}_i, \sigma^2_{\alpha_i})(|\mu_{i\beta}| + K\nu_i)^{-1},$$

where $\hat{\alpha}_i = \sum_{t=1}^{T-1}\bar{\beta}_{i,t+1}\bar{\beta}_{it}/\sum_{t=2}^{T-1}\bar{\beta}^2_{it}$ and $\sigma^2_{\alpha_i} = \sigma^2_{i,\beta}/\sum_{t=2}^{T-1}\bar{\beta}^2_{it}$ with $\bar{\beta}_{it} = \beta_{it} - \mu_i$.

The candidate drawn from the conditional posterior density is accepted with the probability

$$\min\left\{1, \frac{Beta(\alpha^*_i)(1-\alpha^{*2}_i)^{1/2}\{|\mu_{i\beta}| + K\nu^*_i\}}{Beta(\alpha_{i\beta})(1-\alpha^2_{i\beta})^{1/2}\{|\mu_{i\beta}| + K\nu_i\}}\right\}.$$

The conditional posterior density of $\sigma^{-2}_{i,\beta}$

$$p(\sigma^{-2}_{i,\beta}|\mu_{i\beta}, \alpha_{i\beta}, \beta_{i,1:T}, \delta_i) \propto \mathcal{TG}_{D_i}(\sigma^{-2}_{i,\beta}|\hat{v}_i/2, \hat{V}_i/2)(|\mu_{i\beta}| + K\nu_i)^{-1}$$

where the $\mathcal{TG}_{D_i}$ is the density of the implied gamma distribution truncated on D_i, $\hat{v}_i = v_{0i} + T$ and $\hat{V}_i = V_{0i} + (1-\alpha^2_{i\beta})\bar{\beta}^2_{i1} + \sum_{t=1}^{T-1}(\bar{\beta}_{i,t+1} - \alpha_{i\beta}\bar{\beta}_{it})^2$.

Accepting the candidate, drawn from the conditional posterior density, with probability

$$\min\left\{1, \frac{|\mu_{i\beta}| + K\nu_i}{|\mu_{i\beta}| + K\nu^*_i}\right\}.$$

Initial Values: We need to choose initial values for some processes to start the Markov chain. Moreover, the Monte Carlo estimation results should be robust to different initial values. In this regard, we have tested the analysis against different initial values. The results are not intensely different. However it is worthwhile to note that there are some 'bad' initial values. The chains produced by these construct non-positive-definite covariance matrix estimates. In this case, the chain cannot proceed. Yet, once we avoid these initial values, our estimation is robust to different initial values.

For the factors, we choose the principal component analysis estimates as initial values. For other processes $\beta_{0:T}, c_{0:T}, h_{0:T}$, the initial values are drawn from the corresponding unconditional distributions. For instance, $h_t \sim \mathcal{N}(\mu_h, \sigma_h^2/(1-\alpha_h^2))$.

Next, we outline briefly the steps of the MCMC estimation. Note that in each step, updated variables from the previous steps are used.

Step 1: Draw $\beta_{0:T}$
Conditional on $\{\theta_\beta, \delta, c_{1:T}, h_{1:T}, f_{1:T}, \gamma, Y_{1:T}, X_{1:T}\}$, we draw $\beta_{0:T}$ by MH algorithm as explained above, where the candidate is drawn from a no-threshold model distribution.

Step 2: Draw δ
Conditional on $\{\theta_\beta, \beta_{1:T}, c_{1:T}, h_{1:T}, f_{1:T}, \gamma, Y_{1:T}, X_{1:T}\}$, we draw the threshold δ. The candidate is drawn from the conditional prior.

Step 3: Draw $\theta_\beta = \{\mu_\beta, \alpha_\beta, V_\beta\}$
Conditional on $\{\beta_{1:T}, \delta\}$, estimation of θ_β is performed as in a typical AR(1) process. The only difference is that the estimated parameters need to be consistent with the threshold set D_i.

Step 4: Draw $c_{0:T}$
Conditional on $\{\theta_c, \delta, \beta_{1:T}, h_{1:T}, f_{1:T}, \gamma, Y_{1:T}, X_{1:T}\}$, the model can be written easily in a state representation.

$$X_t = c_t + \Lambda_t f_t + \gamma Y_t + e_t$$

$$c_t = \mu_c + \alpha_c(c_{t-1} - \mu_c) + v_{ct}$$

Then the process $c_{0:T}$ is drawn in a forward filtering backwards sampling algorithm (Carter and Kohn (1994)).

Step 5: Draw $\theta_c = \{\mu_c, \alpha_c, V_c\}$

Conditional on $c_{0:T}$, we draw θ_c in a simple AR(1) model.

Step 6: Draw $h_{0:T}$

Conditional on $\{\theta_h, \delta, \beta_{1:T}, c_{1:T}, f_{1:T}, \gamma, Y_{1:T}, X_{1:T}\}$, the stochastic volatility $h_{0:T}$ is drawn in a typical SV estimation method. We use MH algorithm step to accept/reject a candidate drawn from the conditional posterior.

Step 7: Draw $\theta_h = \{\mu_h, \alpha_h, V_h\}$

Conditional on $h_{0:T}$, we draw θ_h in a simple AR(1) model as in Step 5.

Step 8: Draw $f_{1:T}$

Conditional on $\{\delta, \beta_{1:T}, c_{1:T}, h_{1:T}, \gamma, Y_{1:T}, X_{1:T}\}$, the latent factors can be drawn in a similar way as $c_{0:T}$ is drawn in Step 4. To transform the model into state space representation, we need to first transform the factors and Y_t into companion form.

Let $F_t = (f_t', Y_t')'$ be $(K \times 1)$ where $K = k + m$, $\tilde{F}_t = (F_t', \ldots, F_{t-q+1}')'$ be $(Kq \times 1)$, $\tilde{\Lambda}_t = [\Lambda_t, \gamma, 0_{(N\times(Kq-K))}]$ be $(N \times Kq)$, $\tilde{\varepsilon}_t = (\varepsilon_t', 0_{(Kq-K)\times 1}')'$ be $(Kq \times 1)$, and $(Kq \times Kq)$ matrix $\mathbf{\Phi}$ is the companion form of the VAR(q) matrices $\Phi(L)$. Then the state space representation of the factors together with the policy variables is as follows.

$$
\begin{aligned}
X_t &= c_t + \tilde{\Lambda}_t \tilde{F}_t + e_t \\
\tilde{F}_t &= \mathbf{\Phi} \tilde{F}_{t-1} + \tilde{\varepsilon}_t
\end{aligned}
$$

Note that the covariance matrix of $\tilde{\varepsilon}_t$ is degenerate, therefore we need to adjust the Kalman filter accordingly and take the corresponding the first $(K \times 1)$ part of the final draw.

Step 9: Draw $\mathbf{\Phi}, \Sigma$

Conditional on $\{f_{1:T}, Y_{1:T}\}$, estimation of $\mathbf{\Phi}$ and Σ is done as in a typical VAR(1) setting $\tilde{F}_t = \mathbf{\Phi}\tilde{F}_{t-1} + \tilde{\varepsilon}_t$.

Step 10: Draw γ

Conditional on $\{\delta, \beta_{1:T}, c_{1:T}, h_{1:T}, f_{1:T}, Y_{1:T}, X_{1:T}\}$, drawing γ is like drawing a coefficient in a simple linear regression: $X_t - \Lambda_t f_t - c_t = \gamma Y_t + e_t$.

2.B Different Restrictions on Λ_t

The results presented in Table 2.3 are obtained when 7 credit variables are placed at the end of the data set. Hence each of them are forced to be loaded only by one factor. Given these identification restrictions, our model leads us to the above mentioned interpretation of the factors. Regarding this matter, can we improve the results somehow by changing the restrictions in the loadings? The answer is *'not necessarily'*. The zero restrictions in the loading matrix fix the rotation of the factors. Even though we assign new restrictions inspired by the results above (e.g. restricting an unemployment variable to be loaded only by the first factor, an expectation variable to be loaded by only the fifth factor etc.), imposing different restrictions changes the rotation of the factors, thereby changing the meanings of the factors.

Table 2.4 below presents the results when we impose new restrictions on Λ_t. These new restrictions are imposed according to the results in Table 3. As one can easily see, the names and the importance of the factors change dramatically. Now, there is a very distinct 'Hous' factor. The fourth factor now loads on both production and employment variables. The meaning of the fifth factor does not change, it can still be called as expectation factor. Unlike the results of Table 2.3, here finance and inflation factors can be differentiated. Again, one factor, f_2, cannot not explain any significant part of any variable; and one factor, f_7 is uninterpretable as it does not load a particular category.

Table 2.4. Survival Rates of the Factor Loadings under Different Restrictions

	$\mathbf{f_1}$	$\mathbf{f_2}$	$\mathbf{f_3}$	$\mathbf{f_4}$	$\mathbf{f_5}$	$\mathbf{f_6}$	$\mathbf{f_7}$
	Hous	—	*Fin*	*PrEm*	*Expc*	*Inf*	—
Production	0.53	0.15	0.42	**0.78**	0.41	0.43	0.38
(Un)Employment	0.49	0.16	0.24	**0.72**	0.50	0.20	0.46
Housing	**0.93**	0.34	0.26	0.16	0.27	0.27	0.34
Interest Rate	0.33	0.04	0.26	0.38	0.20	0.26	0.26
Inflation	0.23	0.33	0.36	0.20	0.36	**0.45**	0.30
Finance	0.24	0.12	**0.69**	0.15	0.27	0.45	0.24
Money	0.13	0.15	0.34	0.07	0.19	0.33	0.25
Credit	0.46	0.24	0.45	0.25	0.12	0.13	0.37
Expectations	0.53	0.34	0.26	**0.77**	**0.76**	0.01	0.16

2.C Impulse response functions

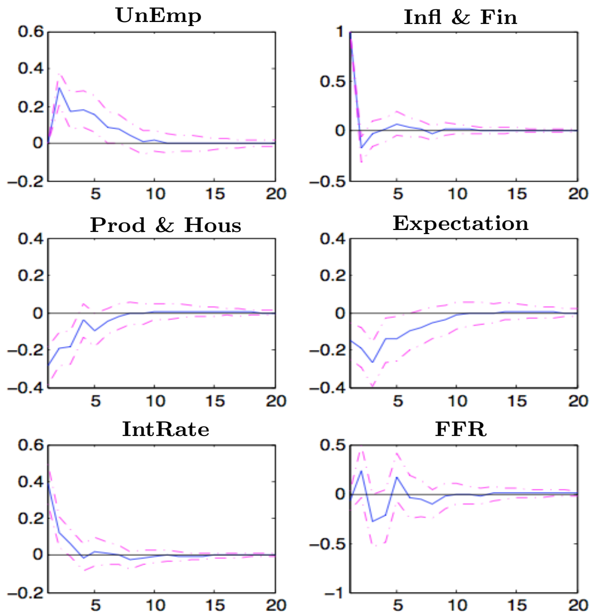


Figure 2.5. The responses of the factors and FFR to a 1 unit shock on *inflation and finance* factor.

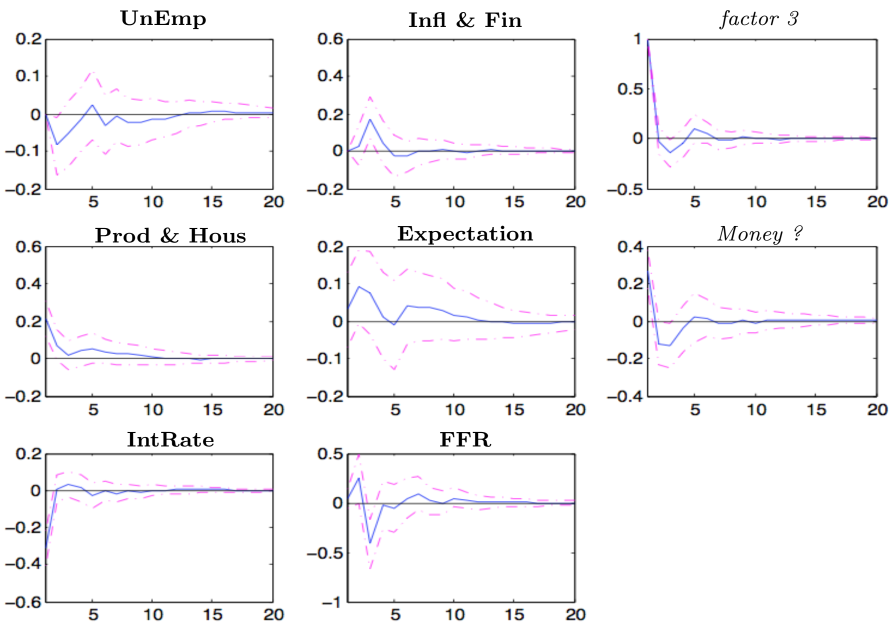


Figure 2.6. The responses of the factors and FFR to a 1 unit shock on *third* factor.

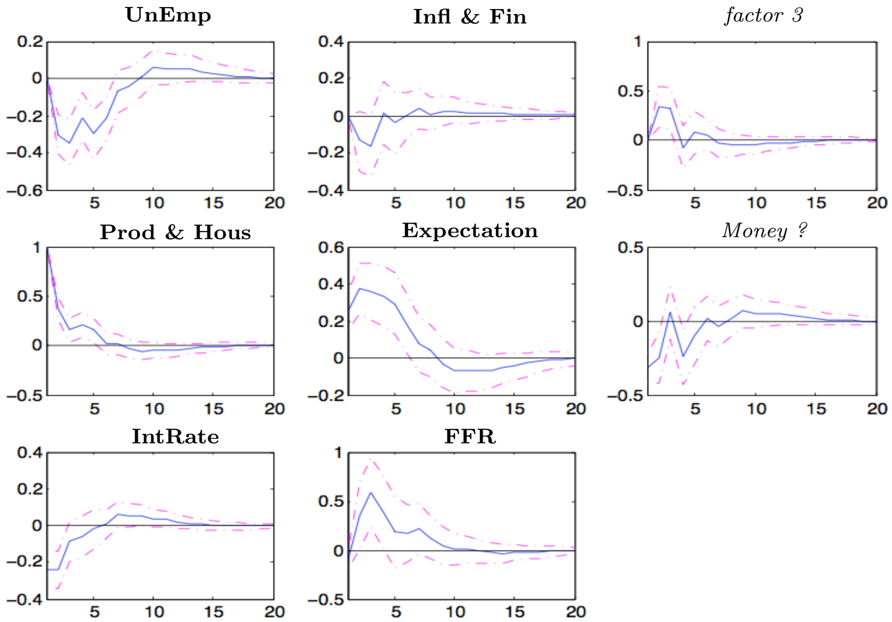


Figure 2.7. The responses of the factors and FFR to a 1 unit shock on *real activity* factor.

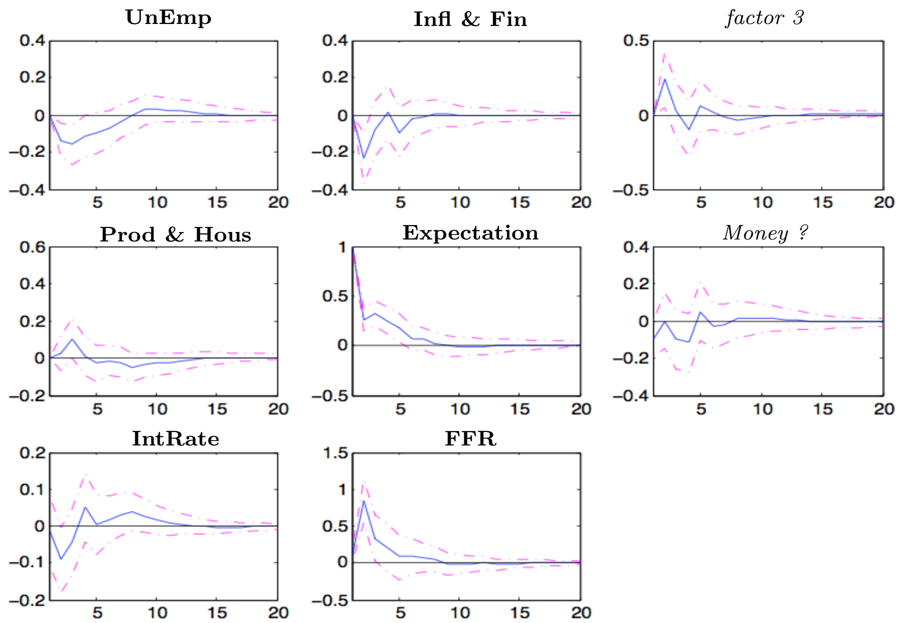


Figure 2.8. The responses of the factors and FFR to a 1 unit shock on *expectations* factor.

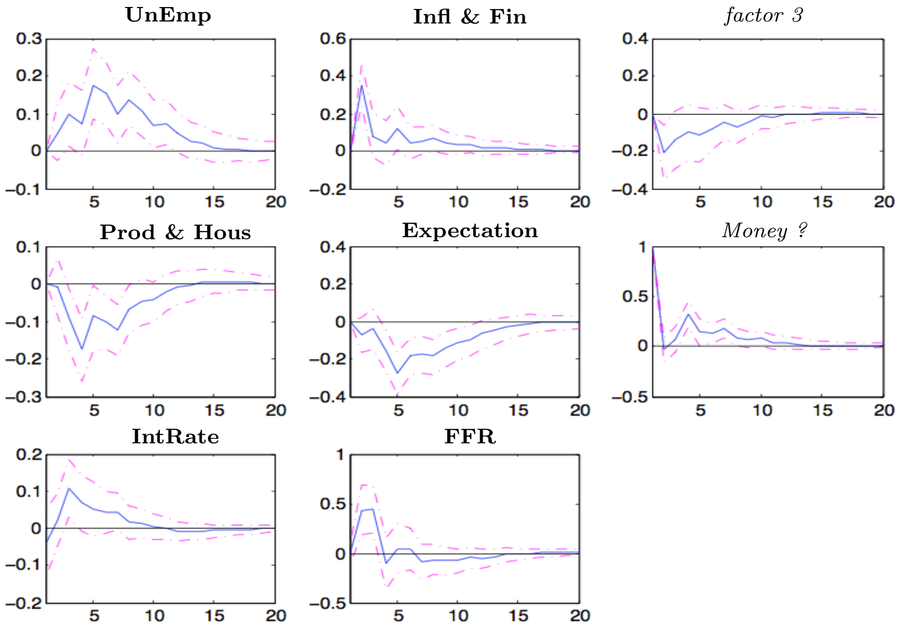


Figure 2.9. The responses of the factors and FFR to a 1 unit shock on *money* factor.

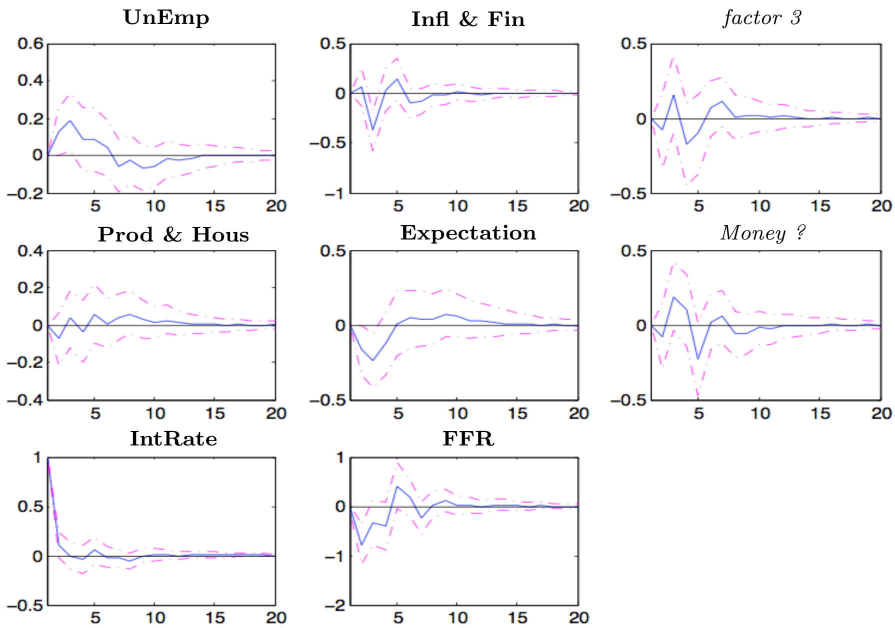


Figure 2.10. The responses of the factors and FFR to a 1 unit shock on *interest rate* factor.

2.D Data Description For SW Updated Data

Series are from generally from FRED. The series indicated whose source is indicated as FRED+SW are mainly gathered from FRED however the missing time periods are patched from Stock and Watson data set. There are two stock exchange variables taken from Shiller's data set, used in Stock Market Data Used in "Irrational Exuberance" Princeton University Press, 2000, 2005, updated. Moreover there are two stock exchange series taken from Stock and Watson data set but patched from www.multpl.com' for the missing values for the last months of the time period. Oil price is included in the data set however has not been used for the analysis in this paper. The analysis requires all series to be stationary. This is ensured by generally by taking differences or logarithms (and in some cases both). The rates are transformed either by keeping them as they are or taking the first or second differences. Similarly the levels are transformed by either taking logarithms or the first or second differences of logarithms. in this respect, 1: levels, 2: first difference, 3: second difference, 4: logarithm, 5: first difference of logarithm, 6: second difference of logarithm.

			Series ID	Tcode	Description	Units	Seasonal Adjustment	Source
1	REAL ACTIVITY	INC	DDURRG3M086SBEA	6	Personal consumption expenditures: Durable goods (chain-type price index)	Index 09=100	SA	FRED
2		INC	DNDGRG3M086SBEA	5	Personal consumption expenditures: Nondurable goods (chain-type price index)	Index 09=100	SA	FRED
3		CONS	DPCERA3M086SBEA	5	Real personal consumption expenditures (chain-type quantity index)	Index 09=100	SA	FRED
4		CONS	DSERRG3M086SBEA	6	Personal consumption expenditures: Services (chain-type price index)	Index 09=100	SA	FRED
5		CONS	PCEPI	6	Personal Consumption Expenditures: Chain-type Price Index	Index 09=100	SA	FRED
6		INC	RPI	5	Real Personal Income	Bil. of Chained 09 $	SAAR	FRED
7		INC	W875RX1	5	Real personal income excluding current transfer receipts	Bil. of Chained 09 $	SAAR	FRED
8		IND	INDPRO	5	Industrial Production Index	Index 07=100	SA	FRED
9		IND	IPFINAL	5	Industrial Production: Final Products (Market Group)	Index 07=100	SA	FRED
10		IND	IPCONGD	5	Industrial Production: Consumer Goods	Index 07=100	SA	FRED
11		IND	IPDCONGD	5	Industrial Production: Durable Consumer Goods	Index 07=100	SA	FRED
12		IND	IPNCONGD	5	Industrial Production: Nondurable Consumer Goods	Index 07=100	SA	FRED
13		IND	IPBUSEQ	5	Industrial Production: Business Equipment	Index 07=100	SA	FRED
14		IND	IPMAT	5	Industrial Production: Materials	Index 07=100	SA	FRED
15		IND	IPDMAT	5	Industrial Production: Durable Materials	Index 07=100	SA	FRED
16		IND	IPNMAT	5	Industrial Production: nondurable Materials	Index 07=100	SA	FRED
17		IND	IPFPNSS	5	Industrial Production: Final Products and Nonindustrial Supplies	Index 07=100	SA	FRED
18		IND	IPFUELN	5	Industrial Production: Fuels	Index 07=100	NSA	FRED
19		UTIL	TCU	1	Capacity Utilization: Total Industry	% of Capacity	SA	FRED+SW
20		UTIL	MCUMFN	1	Capacity Utilization: Manufacturing (NAICS)	% of Capacity	SA	FRED+SW
21	EMPLOYMENT	EMP	CLF16OV	5	Civilian Labor Force	Thous. of Persons	SA	FRED
22		EMP	CE16OV	5	Civilian Employment	Thous. of Persons	SA	FRED
23		UNEMP	UNRATE	2	Civilian Unemployment Rate	%	SA	FRED
24		UNEMP	UEMPMEAN	2	Average (Mean) Duration of Unemployment	Weeks	SA	FRED
25		UNEMP	UEMPLT5	5	Civilians Unemployed - Less Than 5 Weeks	Thous. of Persons	SA	FRED
26		UNEMP	UEMP5TO14	5	Civilians Unemployed for 5-14 Weeks	Thous. of Persons	SA	FRED
27		UNEMP	UEMP15OV	5	Civilians Unemployed - 15 Weeks & Over	Thous. of Persons	SA	FRED
28		UNEMP	UEMP15T26	5	Civilians Unemployed for 15-26 Weeks	Thous. of Persons	SA	FRED
29		UNEMP	UEMP27OV	5	Civilians Unemployed for 27 Weeks and Over	Thous. of Persons	SA	FRED
30		EMP	PAYEMS	5	All Employees: Total nonfarm	Thous. of Persons	SA	FRED
31		EMP	USPRIV	5	All Employees: Total Private Industries	Thous. of Persons	SA	FRED
32		EMP	CES1021000001	5	All Employees: Mining and Logging: Mining	Thous. of Persons	SA	FRED
33		EMP	USCONS	5	All Employees: Construction	Thous. of Persons	SA	FRED
34		EMP	MANEMP	5	All Employees: Manufacturing	Thous. of Persons	SA	FRED
35		EMP	DMANEMP	5	All Employees: Durable goods	Thous. of Persons	SA	FRED
36		EMP	NDMANEMP	5	All Employees: Nondurable goods	Thous. of Persons	SA	FRED
37		EMP	SRVPRD	5	All Employees: Service-Providing Industries	Thous. of Persons	SA	FRED
38		EMP	USTPU	5	All Employees: Trade, Transportation & Utilities	Thous. of Persons	SA	FRED
39		EMP	USWTRADE	5	All Employees: Wholesale Trade	Thous. of Persons	SA	FRED
40		EMP	USTRADE	5	All Employees: Retail Trade	Thous. of Persons	SA	FRED

			Series ID	Tcode	Description	Units	Seasonal Adjustment	Source
41	EMPLOYMENT	EMP	USFIRE	5	All Employees: Financial Activities	Thous. of Persons	SA	FRED
42		EMP	USGOVT	5	All Employees: Government	Thous. of Persons	SA	FRED
43		EMP	CES0000000010	5	Women Employees: Total Nonfarm	Thous. of Persons	SA	FRED
44		EMP	CES0600000007	1	Average Weekly Hours of Production and Nonsupervisory Employees: Goods-Producing	Hours	SA	FRED
45		EMP	AWOTMAN	2	Average Weekly Overtime Hours of Production and Nonsupervisory Employees: Manufacturing	Hours	SA	FRED
46		EMP	AWHMAN	1	Average Weekly Hours of Production and Nonsupervisory Employees: Manufacturing	Hours	SA	FRED
47		EMP	AWHI	5	Index of Aggregate Weekly Hours: Production and Nonsupervisory Employees: Total Private Industries	Index 02=100	SA	FRED
48	HOUSING	HOUS	HOUST	4	Housing Starts: Total: New Privately Owned Housing Units Started	Thous. of Units	SAAR	FRED
49		HOUS	HOUSTNE	5	Housing Starts in Northeast Census Region	Thous. of Units	SAAR	FRED
50		HOUS	HOUSTMW	5	Housing Starts in Midwest Census Region	Thous. of Units	SAAR	FRED
51		HOUS	HOUSTS	5	Housing Starts in South Census Region	Thous. of Units	SAAR	FRED
52		HOUS	HOUSTW	4	Housing Starts in West Census Region	Thous. of Units	SAAR	FRED
53		HOUS	PERMIT	4	New Private Housing Units Authorized by Building Permits	Thous. of Units	SAAR	FRED
54		HOUS	PERMITNE	4	New Private Housing Units Authorized by Building Permits in the Northeast Census Region	Thous. of Units	SAAR	FRED
55		HOUS	PERMITMW	5	New Private Housing Units Authorized by Building Permits in the Midwest Census Region	Thous. of Units	SAAR	FRED
56		HOUS	PERMITS	4	New Private Housing Units Authorized by Building Permits in the South Census Region	Thous. of Units	SAAR	FRED
57		HOUS	PERMITW	4	New Private Housing Units Authorized by Building Permits in the West Census Region	Thous. of Units	SAAR	FRED
58		HOUS	PERMIT1	5	New Private Housing Units Authorized by Building Permits - In Structures with 1 Unit	Thous. of Units	SAAR	FRED
59		HOUS	HOUST1F	5	Privately Owned Housing Starts: 1-Unit Structures	Thous. of Units	SAAR	FRED
60	INTEREST RATE	HOUS	MSACSR	5	Monthly Supply of Homes in the United States	Months' Supply	SA	FRED
61		BILL	CPF3M	2	3-Month AA Financial Commercial Paper Rate	%	NSA	FRED+SW
62		BILL	TB3MS	2	3-Month Treasury Bill: Secondary Market Rate	%	NSA	FRED
63		BILL	TB6MS	2	6-Month Treasury Bill: Secondary Market Rate	%	NSA	FRED
64		BOND	GS1	2	1-Year Treasury Constant Maturity Rate	%	NSA	FRED
65		BOND	DGS3	2	3-Year Treasury Constant Maturity Rate	Avrg.	NSA	FRED
66		BOND	GS5	2	5-Year Treasury Constant Maturity Rate	%	NSA	FRED
67		BOND	GS10	2	10-Year Treasury Constant Maturity Rate	%	NSA	FRED
68		BOND	AAA	2	Moody's Seasoned Aaa Corporate Bond Yield	%	NSA	FRED
69		BOND	BAA	2	Moody's Seasoned Baa Corporate Bond Yield	%	NSA	FRED
70		SPRD	T1YFF	1	1-Year Treasury Constant Maturity Minus Federal Funds Rate	%	NSA	FRED+SW
71		SPRD	T5YFF	1	5-Year Treasury Constant Maturity Minus Federal Funds Rate	%	NSA	FRED
72		SPRD	T10YFF	1	10-Year Treasury Constant Maturity Minus Federal Funds Rate	%	NSA	FRED+SW
73		INTR	INTDSRUSM193N	2	Interest Rates, Discount Rate for United States	% per Annum	NSA	FRED
74		INTR	MPRIME	2	Bank Prime Loan Rate	%	NSA	FRED
75		INTR	INTGSBUSM193N	2	Interest Rates, Government Securities, Government Bonds for United States	% per Annum	NSA	FRED
76	INFLATION	PPI	PPIFGS	6	Producer Price Index: Finished Goods	Index 82=100	SA	FRED
77		PPI	PPIFCG	5	Producer Price Index: Finished Consumer Goods	Index 82=100	SA	FRED
78		PPI	PPIITM	5	Producer Price Index: Intermediate Materials: Supplies & Components	Index 82=100	SA	FRED
79		PPI	PPICMM	5	Producer Price Index: Commodities: Metals and metal products: Primary nonferrous metals	Index 82=100	NSA	FRED
80		PPI	PFCGEF	5	Producer Price Index: Finished Consumer Goods Excluding Foods	Index 82=100	SA	FRED

			Series ID	Tcode	Description	Units	Seasonal Adjustment	Source
81	INFLATION	PPI	PPIACO	5	Producer Price Index: All Commodities	Index 82=100	NSA	FRED
82		PPI	PPICPE	5	Producer Price Index: Finished Goods: Capital Equipment	Index 82=100	SA	FRED
83		PPI	PPICRM	5	Producer Price Index: Crude Materials for Further Processing	Index 82=100	SA	FRED
84		PPI	PPIENG	5	Producer Price Index: Fuels & Related Products & Power	Index 82=100	NSA	FRED
85		PPI	PPIFCF	5	Producer Price Index: Finished Consumer Foods	Index 82=100	SA	FRED
86		PPI	PPIFGS	6	Producer Price Index: Finished Goods	Index 82=100	SA	FRED
87		PPI	PPIIDC	5	Producer Price Index: Industrial Commodities	Index 82=100	NSA	FRED
88		CPI	CPIAUCSL	6	Consumer Price Index for All Urban Consumers: All Items	Index 82-84=100	SA	FRED
89		CPI	CPIAPPSL	6	Consumer Price Index for All Urban Consumers: Apparel	Index 82-84=100	SA	FRED
90		CPI	CPITRNSL	5	Consumer Price Index for All Urban Consumers: Transportation	Index 82-84=100	SA	FRED
91		CPI	CPIMEDSL	6	Consumer Price Index for All Urban Consumers: Medical Care	Index 82-84=100	SA	FRED
92		CPI	CUSR0000SAC	5	Consumer Price Index for All Urban Consumers: Commodities	Index 82-84=100	SA	FRED
93		CPI	CUUR0000SAD	6	Consumer Price Index for All Urban Consumers: Durables	Index 82-84=100	NSA	FRED
94		CPI	CUSR0000SAS	6	Consumer Price Index for All Urban Consumers: Services	Index 82-84=100	SA	FRED
95		CPI	CPIULFSL	6	Consumer Price Index for All Urban Consumers: All Items Less Food	Index 82-84=100	SA	FRED
96		CPI	CUUR0000SA0L2	6	Consumer Price Index for All Urban Consumers: All items less shelter	Index 82-84=100	NSA	FRED
97		CPI	CUSR0000SA0L5	6	Consumer Price Index for All Urban Consumers: All items less medical care	Index 82-84=100	SA	FRED
98		CPI	CUSR0000SAF11	5	Consumer Price Index for All Urban Consumers: Food at home	Index 82-84=100	SA	FRED
99		CPI	CUUR0000SEFV	6	Consumer Price Index for All Urban Consumers: Food away from home	Index 82-84=100	NSA	FRED
100		EARN	CES0600000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees: Goods-Producing	$ per Hour	SA	FRED
101		EARN	CES2000000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees: Construction	$ per Hour	SA	FRED
102		EARN	CES3000000008	6	Average Hourly Earnings of Production and Nonsupervisory Employees: Manufacturing	$ per Hour	SA	FRED
103		EARN	AHETPI	6	Average Hourly Earnings of Production and Nonsupervisory Employees: Total Private	$ per Hour	SA	FRED
104		EARN	CES0500000030	6	Average Weekly Earnings of Production and Nonsupervisory Employees: Total Private	$ per Week	SA	FRED
105	FINANCIAL MARKET	S&P	SP500	5	S&P 500 Stock Price Index	Index	NSA	FRED
106		S&P	SP600DIV	2	S&P 500 Stock Price Index: Dividend	Percent		Shiller
107		S&P	SP500DIVY	2	S&P 500 Dividend Yield	% per Annum		SW+Multpl
108		S&P	SP500EARN	4	S&P 500 Stock Price Index: Earnings	%	NSA	Shiller
109		S&P	SP500PE	5	S&P 500 Price Earnings Ratio	Ratio		SW+Shiller
110		EXRT	EXSZUS	5	Switzerland / U.S. Foreign Exchange Rate	Swiss Francs per Dollar	NSA	FRED+SW
111		EXRT	EXJPUS	5	Japan / U.S. Foreign Exchange Rate	Japanese Yen per Dollar	NSA	FRED+SW
112		EXRT	EXUSUK	5	U.S. / U.K. Foreign Exchange Rate	Dollar per British Pound	NSA	FRED+SW
113		EXRT	EXCAUS	5	Canada / U.S. Foreign Exchange Rate	Canadian Dollar per One U.S. Dollar	NSA	FRED+SW
114		DOWJ	DJCA	2	Dow Jones Composite Average	Index	NSA	FRED
115		DOWJ	DJIA	2	Dow Jones Industrial Average	Index	NSA	FRED
116		DOWJ	DJTA	2	Dow Jones Transportation Average	Index	NSA	FRED
117		DOWJ	DJUA	2	Dow Jones Utility Average	Index	NSA	FRED
118	MONEY	MS	M1SL	6	M1 Money Stock	Bil. of $	SA	FRED
119		MS	M2SL	6	M2 Money Stock	Bil. of $	SA	FRED
120		MS	M2REAL	6	Real M2 Money Stock	Bil. of 82-83 $	SA	FRED

			Series ID	Tcode	Description	Units	Seasonal Adjustment	Source
121	MONEY	MS	AMBSL	6	St. Louis Adjusted Monetary Base	Bil. of $	SA	FRED
122		DEPO	TOTRESNS	6	Total Reserves of Depository Institutions	Bil. of $	NSA	FRED
123		DEPO	NONBORRES	3	Reserves Of Depository Institutions, Nonborrowed	Mil. of $	NSA	FRED
124		MS	BOGAMBSL	5	Board of Governors Monetary Base, Adjusted for Changes in Reserve Requirements (DISCONTINUED SERIES)	Bil. of $	SA	FRED
125		MS	CURRSL	6	Currency Component of M1	Bil. of $	SA	FRED
126		DEPO	DEMDEPSL	6	Demand Deposits at Commercial Banks	Bil. of $	SA	FRED
127		DEPO	EXCRESNS	5	Excess Reserves of Depository Institutions (DISCONTINUED SERIES)	Bil. of $	NSA	FRED
128		MS	MABMM301USM189S	5	M3 for the United States	Natinonal Currency	SA	FRED
129		MS	MBCURRCIR	6	Monetary Base; Currency In Circulation	Mil. of $	NSA	FRED
130		DEPO	NFORBRES	3	Net Free or Borrowed Reserves of Depository Institutions (DISCONTINUED SERIES)	Bil. of $	NSA	FRED
131		DEPO	REQRESNS	6	Required Reserves of Depository Institutions	Bil. of $	NSA	FRED
132		DEPO	RESBALNS	5	Total Reserve Balances Maintained with Federal Reserve Banks	Bil. of $	NSA	FRED
133		DEPO	SAVINGSL	5	Savings Deposits - Total	Bil. of $	SA	FRED
134		DEPO	STDCBSL	5	Small Time Deposits at Commercial Banks	Bil. of $	SA	FRED
135		DEPO	STDSL	5	Small Time Deposits - Total	Bil. of $	SA	FRED
136		DEPO	SVGCBSL	5	Savings Deposits at Commercial Banks	Bil. of $	SA	FRED
137		DEPO	TCDSL	6	Total Checkable Deposits	Bil. of $	SA	FRED
138		MS	M2MOWN	2	M2 Minus Own Rate	%	NSA	FRED
139		MS	M2MSL	5	M2 Less Small Time Deposits	Bil. of $	SA	FRED
140	CREDIT	CRED	BUSLOANS	5	Commercial and Industrial Loans, All Commercial Banks	Bil. of $	SA	FRED
141		CRED	CONSUMER	5	Consumer Loans at All Commercial Banks	Bil. of $	SA	FRED
142		CRED	OTHSEC	5	Other Securities at All Commercial Banks	Bil. of $	SA	FRED
143		CRED	REALLN	5	Real Estate Loans, All Commercial Banks	Bil. of $	SA	FRED
144		CRED	TOTALSL	5	Total Consumer Credit Owned and Securitized, Outstanding	Bil. of $	SA	FRED
145		CRED	NONREVSL	5	Total Nonrevolving Credit Owned and Securitized, Outstanding	Bil. of $	SA	FRED
146		CRED	INVESTNSA	5	Securities in Bank Credit, All Commercial Banks	Bil. of $	NSA	FRED
147		CRED	LOANINVNSA	5	Bank Credit, All Commercial Banks	Bil. of $	NSA	FRED
148		CRED	LOANS	5	Loans and Leases in Bank Credit, All Commercial Banks	Bil. of $	SA	FRED
149		CRED	OLLACBM027NBOG	5	Other Loans and Leases, All Commercial Banks	Bil. of $	NSA	FRED
150		CRED	USGSEC	5	Treasury and Agency Securities at All Commercial Banks	Bil. of $	SA	FRED
151	EXPECTATIONS	PRIX	NAPMPI	1	ISM Manufacturing: Production Index	Index	SA	FRED
152		EMIX	NAPMEI	1	ISM Manufacturing: Employment Index	Index	SA	FRED
153		NEWO	NAPMNOI	1	ISM Manufacturing: New Orders Index	Index	SA	FRED
155		VEND	NAPMSDI	1	ISM Manufacturing: Supplier Deliveries Index	Index	SA	FRED
155		INVT	NAPMII	1	ISM Manufacturing: Inventories Index	Index	NSA	FRED
156		COMM	NAPMPRI	1	ISM Manufacturing: Prices Index	Index	NSA	FRED
157		PURC	NAPM	1	ISM Manufacturing: PMI Composite Index	Index	SA	FRED
158		FFR	FEDFUNDS	2	Effective Federal Funds Rate	%	NSA	FRED
		OIL	OILPRICE	5	Spot Oil Price: West Texas Intermediate (DISCONTINUED SERIES)	$ per Barrel	NSA	FRED

CHAPTER 3

Macroeconomic Tail Events with Non-linear BVARs

3.1 Introduction

This paper seeks to enhance the understanding of macroeconomic tail events. We define macroeconomic tail events as high-impact economic outcomes which arise with a small probability. We adopt a regime switching modeling approach to capture potential nonlinearities in the data and consequently study tail events in the macroeconomy. Using regime switching vector autoregressive (VAR) models, in the spirit of Alessandri and Mumtaz (2014) and Hubrich and Tetlow (2015), we exploit changes in economic dynamics during stressful times,. These non-linear VARs enable us to assign proper economic meanings to different states of the world and to perform structural analyses.

Economic dynamics during economically or financially stressful times are potentially different from normal times. Of particular importance are the nonlinearities induced by the switches in economic regimes and the existence of adverse feedback loops between real and financial sectors, as highlighted by Brunnermeier and Sannikov (2014). Various empirical papers also discuss the importance of nonlinearities, including Hamilton (1989), Kim and Nelson (1999a), Piger et al. (2005), Primiceri (2005), Mishkin (2010). In such situations, linear models lose appeal. As

Drehmann et al. (2007) discusses, linear approximations might sufficiently work in the middle of the distributions but can behave badly in the tails. Consequently, linear models are not adequate for studying tail events and for capturing the possible impacts of adverse shocks. Our adoption of regime switching models explicitly addresses this issue.

We build on Hubrich and Tetlow (2015)'s seminal findings to explore different dynamics of the economy in different states. Apart from Markov switching models, we additionally consider threshold VAR models to study non-linear dynamics. By employing generalized impulse response function analysis, we demonstrate the powerful feedback loops between financial and real sectors, which would otherwise be missed with the conditionally linear impulse responses commonly used in the literature such as by Hubrich and Tetlow (2015), among others. We go one step further to show that out-of-sample *conditional* forecasting exercises performed with threshold VARs carry big potential in forecasting macroeconomic tail events.

We construct a simple system of five variables summarising real economic activity along with its linkages to, and the linkages between, the banking sector and financial markets in the UK. We estimate a linear Bayesian VAR (henceforth BVAR) model as a benchmark, and Threshold VAR (TVAR) models and Markov switching VAR (MSVAR) model.

We consider two different threshold variables in the TVAR exercise: real GDP growth rate and aggregate corporate bond spreads. The use of the real GDP growth rate as a threshold variable is motivated by the desire to study *macroeconomic* recessions. In our case, the estimated threshold value hovers around zero percent, consistent with the common definition of recessions. Hence we can discern *recessionary regimes* from *non-recessionary* ones. The use of corporate bond spreads as the threshold variable is motivated by the observation that not all recessions are related to financial stress, and so our desire to study regimes which are *financially stressful*.

As for the MSVAR, we assume one Markov chain which governs the transition of regimes for both variance and coefficient regimes in a two regime economy. The latent regimes captured by this model are broadly similar to the stressful regimes picked up by our TVARs, on top of the periods characterised by high interest rate and inflation rates in the mid 1980s and early 1990s. Taken together, our regime switching models are able to identify periods of extreme stress in the UK economy.

To study structural shock transmission under different regimes, we adopt the common recursiveness assumption. We study financial shocks (proxied by exogenous jumps in corporate bond spreads), negative output growth shocks and monetary policy shocks, and compute the generalised impulse response functions as described in Koop et al. (1996). The TVAR models successfully capture three valuable facts. First, financial shocks hitting during recessionary periods create disproportionately more severe recessions and significant declines in aggregate bank excess returns. Second, negative output growth shocks occurring during times of financial stress lead to disproportionately higher financial stress. In particular, these shocks in financially stressful periods generate a surge in corporate bond spreads *seven times as large* as growth shocks of a similar size in the financially non-stressful regime. Third, we find some evidence that the drop in output growth induced by interest rate shocks is deeper, but much less persistent in recessions. Aggregate bank excess returns drop in recessions, as opposed to a rise, as predicted by the linear model. Our MSVAR model also generates qualitatively similar results. These results, which the linear BVAR model fails to capture, point out the significance of acknowledging nonlinearities and provide useful information to investigate tail events conditional on structural shocks.

We also provide evidence for the existence of powerful feedback loops between real and financial sectors. Here, we shift our attention to shocks deep in tails (proxied by the size of three-standard-deviation shocks in our exercise) hitting during *normal* times, especially in booms.[1] For example, an extremely large credit spread shock during normal times leads to disproportionately large falls in output, which then feed back into the financial sector resulting in a further increase in financial stress. Such feedback effects can be explained by the endogenous switches of regimes from *non-stressful* to *stressful* ones, which adds to the existing non-linear dynamics in the model.

The above findings also carry important empirical implications for the theoretical results highlighted in Brunnermeier and Sannikov (2014). Their prediction that small shocks can be amplified once the economy is already in crisis regimes is supported by our empirical results that the economic impact is more severe when small shocks hit during recessions or financially stressful times. Our results also speak to their prediction that the economy reacts to large exogenous shocks differently compared to small ones. In particular, our simulations exploiting tail shocks show that the impact on the real and financial sectors are disproportionately larger relative to smaller shocks, suggesting that large shocks are strongly amplified. On the whole, we present strong empirical evidence that reactions to shocks in an economic system can be highly nonlinear.

We then turn to a reduced-form out-of-sample forecasting analysis. Having re-estimated our

[1]This is motivated by the belief that the provability of tail events occurring is time-varying and may depend on financial or business cycles. In the context of supervisory stress-testing, Bank of England (2015) postulates that 'the severity [of scenario] is likely to be greater in a boom, for example, when growth in credit is rapid and asset prices unsustainable high' and hence proposes a counter–cyclical approach in the design of stress scenario.

models with data until 2007:Q2, right before the Great Recession set in, we produce multi–step ahead predictive densities for the variables in the system and compare them against the outturns during the Great Recession period. We seek answers to the following two questions. First, do non-linear models produce *unconditional* density forecasts that are more informative in the tail than their linear counterparts? Second, could we produce more reliable predictive densities when we *condition* on variable paths consistent with stressful times? Notice that we are not trying to predict the timing of crises or any stress events, but rather assessing the capabilities of different models to generate a broad view on the occurrence and nature of tail events.

The answers to the two questions are sequentially 'slightly more' and 'potentially substantial'. We find that unconditional density forecasts produced by the MSVAR and TVAR do not exhibit huge advantages in generating a broad view of tail event forecasts. On the other hand, when we follow Waggoner and Zha (1999) to perform *conditional forecasting* based on variable path of *corporate bond spreads* during the Great Recession, TVAR, and, to a lesser extent MSVAR, show substantial improvements in the tail density forecasts for the output growth. Our findings confirm the importance of incorporating nonlinearities in modelling macro data, and demonstrate the usefulness of such non-linear models to generate reasonable tail forecasts.

There are a range of studies examining tail events/risks. In finance, value-at-risk models are commonly used to measure the loss event on a specific portfolio of financial exposures. This concept has been recently adopted to macroeconomics. Boucher and Maillet (2015) estimate value–at–risk of US output using quantile regressions, and produces a fan chart for out-of-sample forecasts of industrial production growth. This use of quantile regression highlights the importance of outliers, which are associated with extreme events and undoubtedly provide valuable information for modeling and forecasting future tail events. This point is shown in Covas et al. (2014) who use fixed effect quantile autoregressive models to capture the dynamic of banks losses and revenues and to project capital shortfalls, and in Adrian and Brunnermeier (2014) who propose a measure of systemic risk conditional on an institution under stress with quantile regressions. Recently, White et al. (2015) provide a theoretical framework to estimate and make inference in multivariate, multi-quantile models.

While acknowledging the usefulness of quantile regressions, we see two advantages of using non-linear VARs. First, we can explicitly identify different regimes of the economy. Second, we can refrain from making assumptions on which quantiles shocks have to originate in order to contribute to macroeconomic tail events. Our approach fully accommodates scenarios where small shocks can be amplified to create big economic impact.

Our paper is related to an active area of empirical research using non-linear VAR models. Sims and Zha (2006) employ a multivariate regime switching model for US monetary policy in a structural VAR framework. Alessandri and Mumtaz (2014) construct a set of linear and non-linear econometric models to study predictive densities, especially by focusing on tails to assess the power of financial indicators for output and inflation in the US. Last but not the least, Hubrich and Tetlow (2015) investigate the interaction between a financial stress index for the US and real activity, inflation, monetary policy using a Markov switching VAR model. The empirical findings support the inadequacy of single regime models to capture the dynamics of the economy.

The remainder of this paper proceeds as follows. We introduce the econometric models used in this paper in section 3.2. This section also provides details on which priors we use and how they are incorporated to the Gibbs sampler. Section 3.3 presents the features of our data set. We illustrate the estimation results of the proposed procedure in section 3.4. Section 3.5 provides our analysis on the structural shock transmission in our models. Section 3.6 follows with the forecasting results. We conclude with section 3.7 and some other details such as the graphs of individual series are given in the appendix.

3.2 Model Specifications

This section describes the three models we deploy: threshold VAR, the Markov switching VAR and the linear VAR. One major difference between the first two types of models is that in a TVAR, the researcher has to *pre-define* a threshold variable. In other words, any switches of regimes in TVAR are solely determined by the dynamics of the chosen threshold variable. In contrast, regime switches in MSVAR are determined by the joint dynamics of the economic system but the interpretation of such regimes might not be as straightforward.

We give a brief overview of each of the models below, and refer the readers to the appendix and Barnett et al. (2010) for technical details. All models are estimated with two lags.

3.2.1 Threshold VAR

A threshold VAR model comprises an explicit threshold variable which allows regimes to switch endogenously. The associated model is

$$Y_t = \left[c_1 + \sum_{j=1}^{P} \beta_{1,j} Y_{t-j} + v_t\right] R_t + \left[c_2 + \sum_{j=1}^{P} \beta_{2,j} Y_{t-j} + v_t\right] (1 - R_t)$$

where

$$R_t = 1 \iff Z_{t-d} \leq Z^* \quad (\text{or } \; Z_{t-d} \geq Z^* \text{depending on the threshold variable})$$

and $v_t \sim N(0, \Omega_{R_t})$.

The delay parameter, d, is also referred as threshold lag. We consider two different threshold variables: real GDP growth rate and the level of aggregate corporate bond spreads. We define $R_t = 1$ as the *recessionary regime* if and only if the real GDP growth is below an estimated threshold rate for $d = 0$ and $d = 1$ simultaneously, given the common definition of recessions associated with two consecutive periods of negative output growth. Otherwise, the regimes are defined as *non-recessionary*. Our estimated threshold rate hovers around zero percent, which validates this definition. We denote this system as *TVAR-Y*.

The use of corporate bond spreads as the threshold variable is motivated by the observation that not all recessions are related to financial stress, and by our desire to explicitly define financially stressful regimes. We define $R_t = 1$ as the *financially stressful regime* if and only if the credit spreads rise beyond an estimated threshold value for $d = 1$. Otherwise, the regimes are *financially non-stressful*. Our estimated threshold value for the credit spreads is 290 basis points. We denote this system as *TVAR-S*.

To estimate this model, we follow Alessandri and Mumtaz (2014) in using the Gibbs sampling algorithm which includes a Metropolis-Hastings step for sampling the threshold value in each simulation. The threshold Z^* is assumed to have a normal prior, $Z^* \sim N(\bar{Z}, \bar{V})$ where $\bar{Z}$ and $\bar{V}$ are the sample mean and variance of the threshold variable, respectively.

3.2.2 Markov Switching Vector Autoregression Models

The Markov switching VAR Model (MSVAR) is written as

$$Y_t = c_{S_t} + B_{1,S_t}Y_{t-1} + B_{2,S_t}Y_{t-2} + \cdots + B_{l,S_t}Y_{t-L} + v_t, \tag{3.1}$$

where $v_t \sim N(0, \Omega_{S_t})$.

The regime switches follow a joint dynamic for coefficients and variance at the same time. The latent regimes S_t are assigned as $S = 1, 2$. The switch between these latent states is governed by the transition matrix, P :

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix}$$

where $p_{ij} = prob(S_t = i|S_{t-1} = j)$ indicates regime i is followed by regime j. There are no restrictions on regime switches, i.e they are left unrestricted to jump back and forth. The columns sum up to 1.

3.2.3 Benchmark Bayesian VAR

We use a Bayesian VAR model a benchmark against which to compare our non-linear VAR models, and it is given by

$$Y_t = c + B_1Y_{t-1} + B_2Y_{t-2} + \cdots + B_lY_{t-L} + v_t, \tag{3.2}$$

For all models, we impose normal inverse Wishart priors following Bańbura et al. (2010), see the Appendix for more details.

3.3 Data

We use quarterly data from 1965:Q2 to 2014:Q2. We have a 5 variable system including output growth (measured by quarterly change in real GDP), inflation (measured by the quarterly change in consumer price index), excess bank returns, corporate bond spreads (proxied by the difference between UK Corporate Bond Yield and UK 10-year Government Bond Yield) and the short term interest rate (proxied by the three month interest rate). Except for the short term interest rate and corporate bond spreads, all series are annualised.

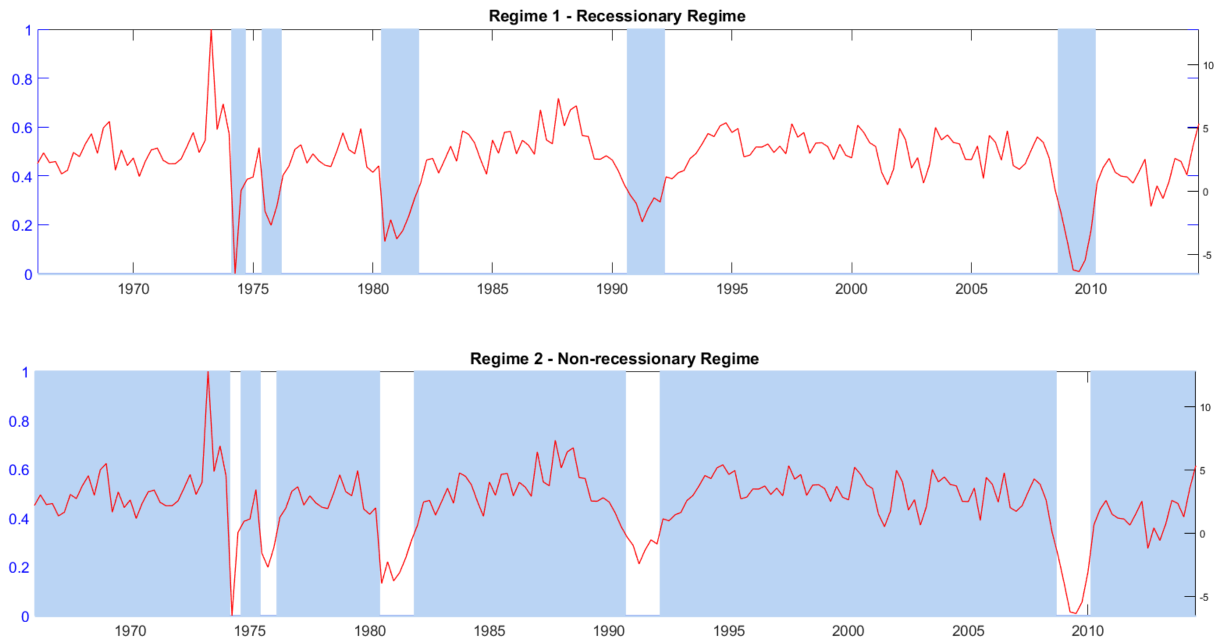
Using corporate bond spreads as a measure of financial frictions follows mainly from Gilchrist and Zakrajšek (2012) and Philippon (2009). The former examines the relationship between corporate bond spreads and economic activity. Excess bond premium, which is one of the two components of GZ index representing the cyclical changes between measured default event and credit spreads, is found to be a predictor of economic activity. Philippon (2009) looks at a similar concept from the market's perspective. It shows that, similar to Tobin's q, a market–based measure of a q can be constructed from corporate bond prices and it performs much better than the traditional one. Furthermore, it is worthwhile to note that the Financial Stress Index constructed by Hubrich and Tetlow (2015) for the US economy includes corporate bond spreads. Our rationale of choosing corporate bond spreads as a proxy of financial stress is, in addition to the use in the literature, for the search of a consistent and long enough series reflecting financial stress. In a similar spirit, excess bank returns are used as a proxy of aggregate bank profitability.

The choice of variables reflects our goal to capture the overall dynamics in the economy and to link the real economic sector to the banking and financial sectors while maintaining a parsimonious model. Tables 3.1 and 3.2 describe the basic statistics of the data alongside the data sources, and Figure 3.2 plots the five variables of interest.

3.4 Full sample estimation results

This section reports the estimation results for both non-linear models using the full sample from 1965:Q2 to 2014:Q2. Figures 3.1 and 3.2 respectively illustrate the recessionary and non-recessionary regimes and financially stressful and non-stressful regimes as modeled by the TVAR-Y and TVAR-S models. Figure 3.3 reports the estimated regime probabilities addressing high and low stress states implied by our MSVAR model.

In a TVAR model, the regime changes are abrupt and the economy is either in one regime or the other. Therefore the probabilities accompanying the regimes are either 0 or 1. The first regime in Figure 3.1 is labelled as *recessionary* whereas the second regime as *non-recessionary* periods. The first two recessions coincide with the mid–1970s recessions. They are associated with the 1973 oil crisis and stagflation that are followed by the decline of traditional British industries and inefficient production caused by excessive union wage demands. The next is the recession in the early 1980s due to deflationary government policies including spending cuts, pursuance of monetarism to reduce inflation, and switches from a manufacturing economy to a services economy. The fourth is the recession in the early 1990s which started in the third quarter of 1990 and went on for five quarters. It was primarily caused by high interest rates, falling house prices and an overvalued exchange rate. Membership of the Exchange Rate Mechanism (1990–1992) was the key factor in keeping interest rates high. The last recession is the Great Recession when the annualised GDP fell almost -7% in the first quarter of 2009.

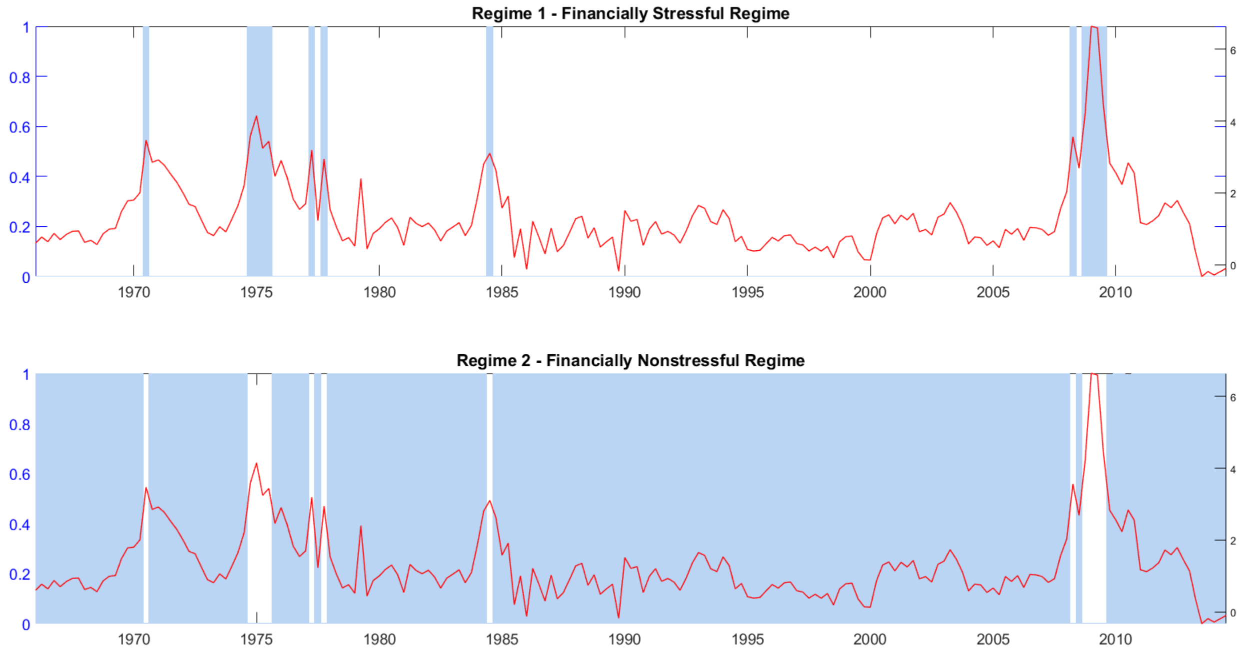


Notes: The threshold variable is the real GDP growth which is overlapped with the regimes.

Figure 3.1. Full sample regimes for TVAR-Y

Similarly, Figure 3.2 presents the regimes where the corporate bond spreads are used as the threshold variable. The first financial stress period corresponds to the second quarter of 1970. The second reflects the impact of the oil crisis and stagflation on the corporate bond market over the periods 1974:Q3–1975:Q2. The third and fourth are associated with 1977:Q1 and 1984:Q2. The former may correspond to the period of introduction of small number of floating–rate issues. The latter is the deregulation of the Eurobond market and opening of international markets. The last financial stress periods show the effects of the Great Recession onto the market and overlap with the recessionary regimes with an exception of the second quarter of 2008.

The regimes given by the MSVAR model in Figure 3.3 suggest a similar interpretation. In



Notes: The threshold variable is the corporate bond spreads series which is overlapped with the regimes.

Figure 3.2. Full sample regimes for TVAR-S

accordance with the data we use (shown in Figure 3.2), the 'high stress regimes' do not only pick up recessionary or financially stressful periods, but also times when the inflation rate and short term interest rates were both high. High stress periods start around 1973 and mute by early 1990s with a significant low stress periods in between 1982 and 1985. Stable periods last for around fifteen years after 1992, followed by the stress periods during the Great Recession.

3.5 Impulse Response Analysis of Structural Shocks

Our setup of different models provides us a convenient platform to study structural shocks based on historical data. We are particularly interested in the potential differences in the shock transmission under specific regimes. This is crucial as the linear models are generally found to be inadequate to explore tail events and to investigate different shock transmission during stressful times.[2]

Broadly, there are two main findings. First, the transmission of shocks in recessionary or finan-

[2]Such a set–up offers insights on the impact of adverse shocks hitting certain *bad* states of the world, which can be potentially useful in calibrating macro scenarios where multiple adverse shocks hit an economy sequentially, for instance, in stress testing.

Figure 3.3. Full sample regime probabilities for MSVAR

cially stressful periods is different than normal times. The distinction between the reaction of the economy in crises and normal times is highlighted Brunnermeier and Sannikov (2014). Our results stress the importance of acknowledging nonlinearities and distinguishing different states of the world. Second, we illustrate Brunnermeier and Sannikov (2014)'s point that the large shocks are strongly amplified by simulating our models with extremely big shocks. We discuss how these tail shocks cause disproportionately deeper stress and materialise feedback loops.

3.5.1 Generalised impulse response functions and shock identification

Koop et al. (1996) discuss that impulse responses in non-linear models are dependent on size, sign and history, which is in contrast to those computed by linear models. They introduce the generalised impulse response functions (GIRFs), which fully takes into account of the possibility of endogenous switches of regimes during simulations.

Following Koop et al. (1996), we calculate these impulse responses as

$$GIRF = E(y^{*,p}_{t+k}|y_t, \Upsilon, \triangle) - E(y^*_{t+k}|y_t, \Upsilon)$$

where k is the forecast horizon, Υ denotes the hyperparameters, $\triangle$ indicates the perturbed shocks while superscript p marks the forecasts with the perturbed path of errors. The appendix reports the steps on the non-linear impulse response functions for both TVAR and MSVAR models.

We adopt the Cholesky decomposition for our purposes, which is common in the empirical macroeconomic literature to identify macro structural shocks. As is well known, the order of variables matters. In our case, we order the variables as: (i) real GDP growth; (ii) inflation rate; (iii) aggregate bank excess returns; (iv) corporate bond spreads; (v) short term interest rate. This order is consistent with the monetary policy literature as real variables respond to monetary policy shocks with a time lag whereas monetary policy responds to shocks from the real sector and the financial sector contemporaneously. Our identification strategy is closely in line with Christiano et al. (1998) and Gilchrist and Zakrajšek (2012).[3]

In this section we consider three structural shocks: (i) corporate bond shocks which proxy for shocks in the financial market; (ii) output growth shocks; (iii) interest rate shocks. We seek to investigate the implications of one-standard-deviation shocks alongside three-standard-deviation shocks.

3.5.2 Shock propagation and macro tail events

Figures 3.4 to 3.13 depict the impulse response functions of all three models to three structural shocks of one-standard-deviation size. We first study *financial shocks*. Gilchrist and Zakrajšek (2012) explain that credit spreads can reflect the changes in the quality of corporate firms' balance sheet and their external finance as well as the capital position of financial intermediaries who supplies credit. Nevertheless, we interpret any exogenous rise in the corporate bond spreads as shocks to the financial intermediation process which is orthogonal to shocks to the real sector. In our linear model given in Figure 3.4, this particular shock leads to a contraction in output growth, an initial drop in bank returns and inflation in the linear model.

The responses of non-linear models, on the other hand, have noticeably different features. Figures 3.5, 3.6 and 3.7 show the impulse responses of of TVAR-Y, TVAR-S, MSVAR, respectively,

[3] We checked alternative orderings of the last three variables. Results are mostly robust except when we order corporate spreads after the interest rate, i.e. when corporate bond spreads react to monetary policy shocks with a time lag but not the other way round. We note that such an ordering is used by Hubrich and Tetlow (2015) where they construct a *monthly* system. We stress that our baseline ordering is in line with Gilchrist and Zakrajšek (2012) who place excess bond premium (a component of credit spreads) before the effective Federal funds rate, and is more convincing because monetary policy is able to respond to financial shocks within the same quarter.

corresponding to a one-standard-deviation shock to corporate bond spreads. In TVAR-Y (Figure 3.5), a one-standard-deviation financial shock in the recessionary world (80 basis point jump in bond spreads) generates a significant and deep decline in output growth with a maximum fall of 50 basis points 5 quarters after the shock. If such shocks happen in the non-recessionary world, the recessionary impact is comparatively much shallower (after factoring in the different size of the shocks). The financial shock also leads to a deeper and more persistent drop in the short-term interest rate in the recessionary world. Interestingly, this shock leads to a rise in inflation rate, which can be explained by the association of the identified recessionary regimes with high inflation rate rates in the 1970s. The most compelling implication is the response of the excess bank returns, a proxy for bank profits. In the recessionary regimes, financial shocks lead to almost 10 pp drop in aggregate bank excess returns two quarters after the shock, whereas there is a slight rise in the non-recessionary world. This reflects that aggregate excess returns of banks are seriously affected when financial shocks hit during recessions.

The GIRFs in TVAR-S, shown in figure 3.6, point a similar picture. A financial shock in an initially financially stressful regime leads to a much greater contraction of output and greater drop in aggregate bank returns when compared to the financially non-stressful world, although the significance is marginal. Similarly, the responses of the MSVAR model in figure 3.7 are qualitatively similar to those of the TVAR-Y and the VAR-S models. Our observation related to excess bank returns rising in the good state also holds for the MSVAR.

These results are consistent with the predictions of the theoretical literature. During times where the real economy is in recession or the financial market is under stress, where balance sheets of the financial and non-financial sectors are weak, any shocks in the financial market further amplify the recessionary effects through the well-known financial accelerator mechanisms as described in Kiyotaki and Moore (1997) and Bernanke et al. (1999). This undoubtedly leads to an erosion of aggregate banks' profits. Our results are also in line with Brunnermeier and Sannikov (2014) whose theoretical model predicts that small shocks can be amplified once the economy is in crisis regimes.

We carry out a similar exercise by hitting in the system with a *negative output growth shock*. Since our identification scheme does not allow us to distinguish aggregate supply shocks from aggregate demand shocks[4], we consider this shock as proxying exogenous changes in the real

[4] Aggregate supply shocks are characterised by a fall in output growth and a rise in inflation rate, whereas aggregate demand shocks are characterised by both decrease in output growth and inflation. These two shocks can be further identified by sign restrictions, which is left for future research.

economy that reduces output growth, which could originate domestically from productivity or demand shocks or internationally through the export-import channel.

Figures 3.9 and 3.10 display the impulse response functions of two threshold VAR variants. As inflation jumps on impact, the models seem to be picking up aggregate supply shocks. Relative to its impact in the non-recessionary/financially non-stressful regimes, this shock leads to a larger and more persistent rise in the inflation rate in the recessionary/financially stressful regimes. This is again heavily driven by the stagflation experience in the UK. The rise in corporate bond spreads is also significant, especially in the financially stressful regimes. In Figure 3.10, the spread can rise up to 22 basis points on impact in the financially stressful regimes as opposed to 3 basis points otherwise, even though the size of growth shocks is similar (about 1.7 pp) across the two states of the world. This shows that when the economy is under financial stress, further bad shocks originating in the real economy lead to heightened stress in the financial system. This point is not picked up by the linear BVAR response as shown in Figure 3.8. GIRFs from the MSVAR model exhibit similar features.

We also consider the generalised impulse response functions of *short term interest rate shocks*, defined as any interest rate movement unexplained by the systematic responses of policy makers to variations in the state of the economy (see Christiano et al. (1998)). Examining the impulse responses of the linear model in Figure 3.12, we observe that the identified interest rate shock leads to a hump–shaped response of real output growth. Corporate spreads rise significantly two years after the shock. These results are consistent with the traditional monetary policy literature.[5] Aggregate bank excess returns go down by 3 percent (annualised), one quarter after the shock.

Figure 3.13 shows the GIRFs of the TVAR-Y model to the interest rate shock under recessionary and non-recessionary regimes. There are several major differences relative to the linear impulse responses. First, the drop in output growth is deeper but much less persistent in recessions. Second, aggregate bank returns drop by about 2 pp (annualised) in recessions as opposed to a rise otherwise. This may be an evidence that banks are particularly vulnerable to monetary policy shocks during recessions. Slightly puzzling is the initial significant drop in corporate bond spreads in recessions.[6]

[5] The inflation rate rises as a result of the shock. This constitutes the *price puzzle*, which is at odds with the theoretical literature which predicts a fall in prices with contractionary monetary policy shocks. A vast literature has discussed the puzzle and proposed solutions, see Sims (1992a), Castelnuovo and Surico (2010), Hürtgen and Cloyne (forthcoming), Christiano et al. (1996) and Balke and Emery (1994).

[6] The TVAR-S model produces a boom conditional on an unexpected rise in short term interest rate. The results seem counterintuitive so we do not report them here.

3.5.3 Tail shocks and the feedback loop

Our non-linear models allow us to investigate shock transmission mechanisms when extremely large shocks hit the economy, especially when the economy is in *normal* times. This is motivated by the belief that the occurrence of tail events is time varying and may depend on financial or business cycles. This point is addressed by Bank of England (2013) in the context of scenario design in stress testing.[7]

We repeat our GIRF simulations with *three-standard-deviations* shocks, which arguably represent tail events. We first consider a tail shock in output growth during non-recessionary regimes in the TVAR-Y model, the results of which are reported in the grey shades in Figure 3.14. When compared with an output growth tail shock of the size 8 pp (annualised) in the recessionary regimes which leads to a maximum rise in credit spreads of 26 bp, a 4 pp shock in the non-recessionary regimes cause as a surge in spreads as high as 30 bps within the first five quarters. Not only the magnitude but also the persistence in the rise in credit spreads becomes very severe. Particularly noticeable is its hump shape response starting from the third quarter. Such response is markedly different from those attributed to the standard one-standard-deviation shock shown in Figure 3.9. Such significant response in financial stress then feeds back to the real sector, giving rise to a far deeper and protracted recession as shown in the GDP growth response in Figure 3.14.

We perform similar simulations with credit spread shocks in the TVAR-S model. A tail shock comprising a 120 bp surge in credit spreads in financially non-stressful regimes, as shown in Figure 3.15, results in a very deep and persistent recession, with the trough occurring three years after the shock. The contraction reaches 1.2 pp, as opposed to the much shallower recession of a maximum contraction of 0.1 pp with the one-standard-deviation shock displayed in Figure 3.6. Again, due to the powerful feedback effects from the real sector, the persistence and rise in corporate spreads in the future horizon are more protracted. These responses are not only disproportionately larger when compared to the one-standard-deviation shock scenario, but also larger relative to the responses in the financially stressful regimes.

These results can be intuitively explained by the simulations which allow endogenous switches

[7] Bank of England (2013) states that one way to explore the severity of the scenarios in stress-testing is 'to recognise the variation in the probability and impact of systemic stresses over time'. For example, as credit conditions ease and leverage builds up, the banking system may be susceptible to more severe shocks. Conversely, in a downturn, with tightening credit conditions and lower leverage, a less severe scenario might be more appropriate. Our set-up also enables us to carry out an experiment in this regard.

of regimes from *non-stressful* to *stressful* when such tail shocks hit, which further amplifies the existing non-linear dynamics. Our empirical findings provide strong support to non-linear general equilibrium models for financial stress as pioneered by Boissay et al. (2013) and Brunnermeier and Sannikov (2014), among others.

3.6 Forecasting analysis

In this section, we investigate how well these models are able to forecast tail events. In particular, we perform *unconditional* and *conditional* forecasting. We are interested in the following two questions. First, do non-linear models produce more informative unconditional density forecasts in the tail? Second, could we produce more reliable predictive densities when we condition on variable paths consistent with stressful times? It is worth emphasizing that we are interested in checking whether our models are able to provide a broad view of plausible macro tail events. We do not seek to predict the exact timing of stresses. Both exercises involve re–estimating the three models using data until 2007:Q2 and produce multi–step ahead, pseudo-out-of-sample forecasts for 12 quarters. We compare the predictive densities against the periods including the Great Recession in 2008–09.

The rationale for performing an unconditional forecasting exercise is to investigate whether the non-linear models are able to generate forecast densities which properly characterize tail events without conditioning on any information. Figure 3.16 reports the fan charts of the output growth and the credit spreads for all three models.[8] The black lines are the median forecasts whereas the shaded areas are the error bands from 20^{th} to 80^{th} percentiles with 5% increments. The charts are overlapped with the realisations of both series which are given by the red lines. Visually speaking, none of these models are capable of capturing the drop in output growth during the 2008-09 Great Recession, a truly extreme event in the post-war sample although there may be some slight improvement in the coverage of tails for MSVAR model.

We then study conditional forecasting to investigate whether our models are better at capturing tail events if we *ex ante* feed into specific paths of variables under stress.[9] This enables the models to exploit the correlational dynamics between the conditioning variable and the other variables

[8]To keep the charts concise, we only report TVAR-Y results for the forecasting exercises.

[9]Conditional forecasting is a common exercise among policy makers. For example, in the Inflation Reports produced by the Bank of England, the fan chart projections for GDP growth and CPI inflation are generated conditional on 'market interest rate expectations' and 'the stock of purchased assets financed by the issuance of central bank reserves'.

being modeled. We employ the conditional forecasting techniques proposed by Waggoner and Zha (1999) and used by Bańbura et al. (2015a). The algorithm first involves sampling the residuals of the VAR system which are consistent with the exogenously imposed scenario paths, and then constructs forecast densities of other variables conditional on the imposed paths. In this exercise, we impose the actual out-turns of corporate bond spreads between 2007:Q3 and 2010:Q2. Figure 3.17 reports the corresponding fan charts.

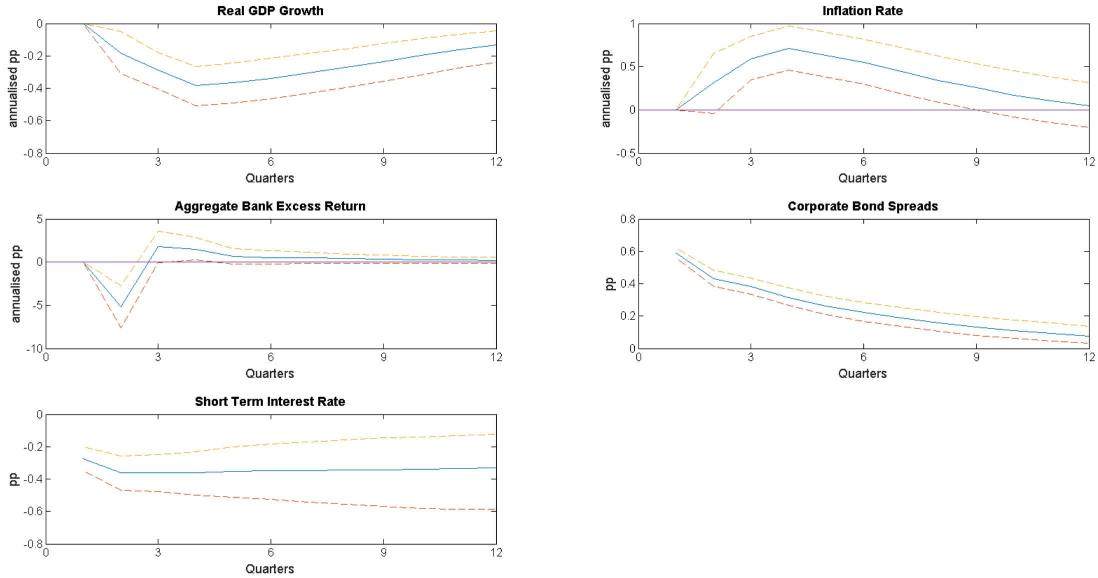
We draw the following conclusions. First, the linear BVAR model consistently falls short of producing reasonable probe of macro tail events, as shown in the first column of Figure 3.17. Second, the TVAR-Y model in the second column shows the most significant improvement in the forecast densities compared to other models. Third, the MSVAR model in the third column shows some improvement in the forecast density of the output growth, although the improvement is not as significant as in TVAR-Y.

These results highlight the importance of incorporating nonlinearities in forecasting macro-financial variables, especially in the conditional forecasting exercise. This appears to support Clements and Smith (2000) that non-linear models can potentially perform better than the linear counterparts in terms of the density forecast precision, as long as the data contain non-linear features.

3.7 Concluding remarks

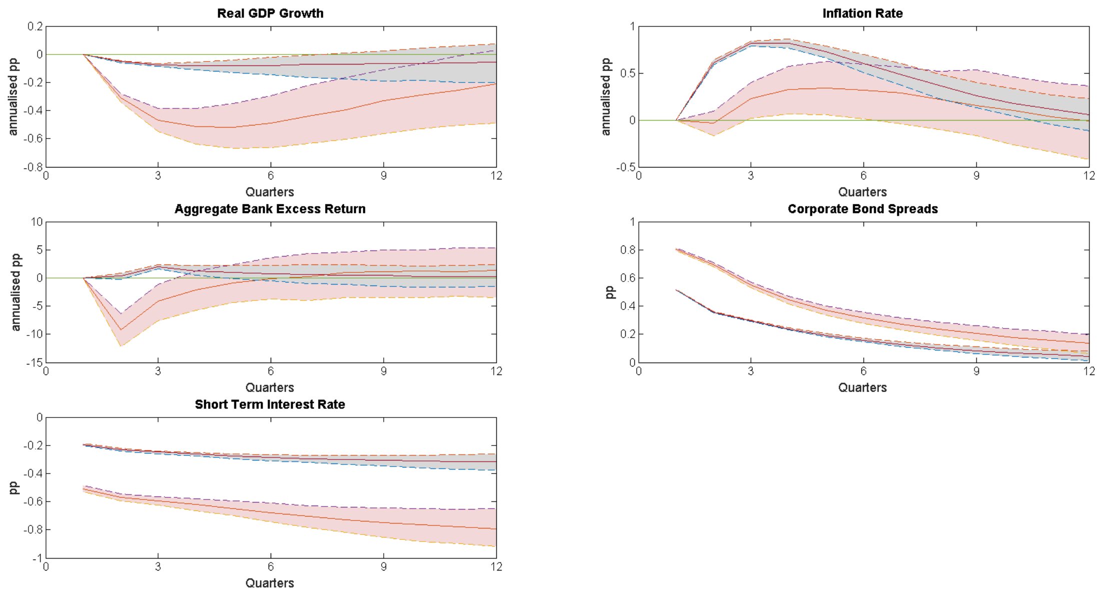
In this paper, we estimate a set of non-linear Bayesian VARs to study macroeconomic tail events. We utilise regime switching models to estimate the regimes governed by different time periods. Our estimated regimes are associated with recessionary/non-recessionary and financially stressful/stable periods. We obtain substantial evidence that financial shocks during recessionary periods cause disproportionately more severe contractions in the real sector. We also demonstrate the existence of a powerful feedback loop between the real and financial sectors as a result of tail shocks hitting the economy in non-recessionary/stable time periods. These findings serve as empirical support to the theoretical predictions in Brunnermeier and Sannikov (2014). Moreover, we check each model's out-of-sample forecasting power, and find that conditional predictive densities produced by TVARs hold the potential to explore downside events.

Future work involves extending our results to the literature of forecasting macroeconomic tail risks, as in Boucher and Maillet (2015) and De Nicolo and Lucchetta (2016), where the probabilities of the occurrence of tail events are also under investigation. Our generalised impulse responses provide the densities of 'variables' at each forecast horizon conditional on a structural shock hitting at a particular economic or financial regime. With additional information on the densities of the 'regimes', conditional distributions of variables can be further transformed to marginal distributions to study tail risks. We leave this for future research.



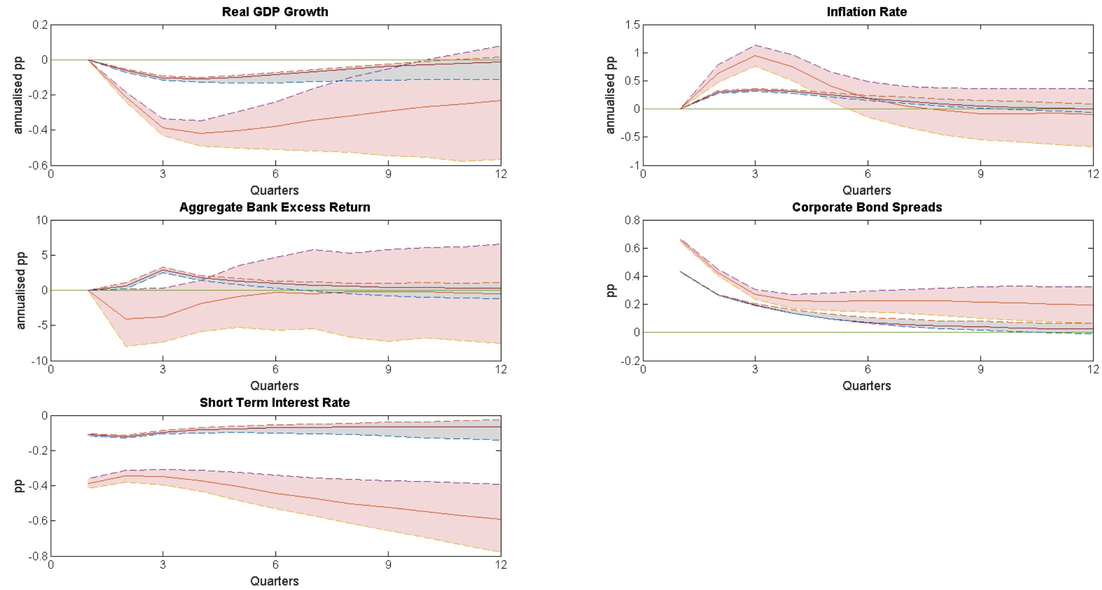
Notes: The error bands correspond to the 68% error bands.

Figure 3.4. Impulse Responses to a 1 SD adverse shock to corporate bond spreads in BVAR model



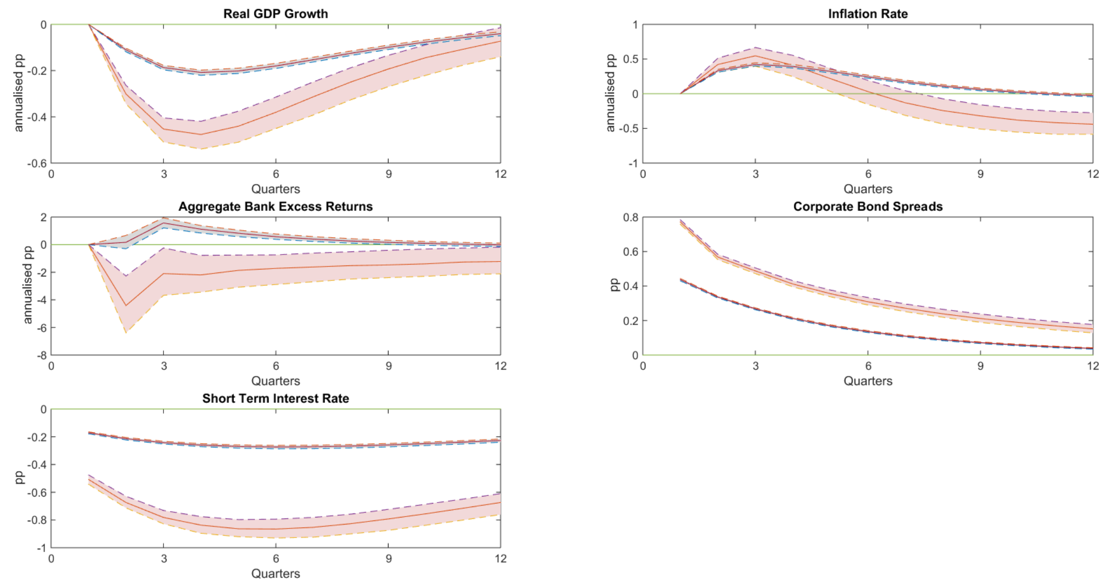
Notes: The error bands correspond to the 68% error bands. The red shaded area corresponds to the error bands of recessionary regimes. The grey shades correspond to those of non-recessionary regimes. The green line is the median response of the BVAR model to the same shock.

Figure 3.5. Generalised Impulse Responses to a 1 SD adverse shock to corporate bond spreads in TVAR-Y model



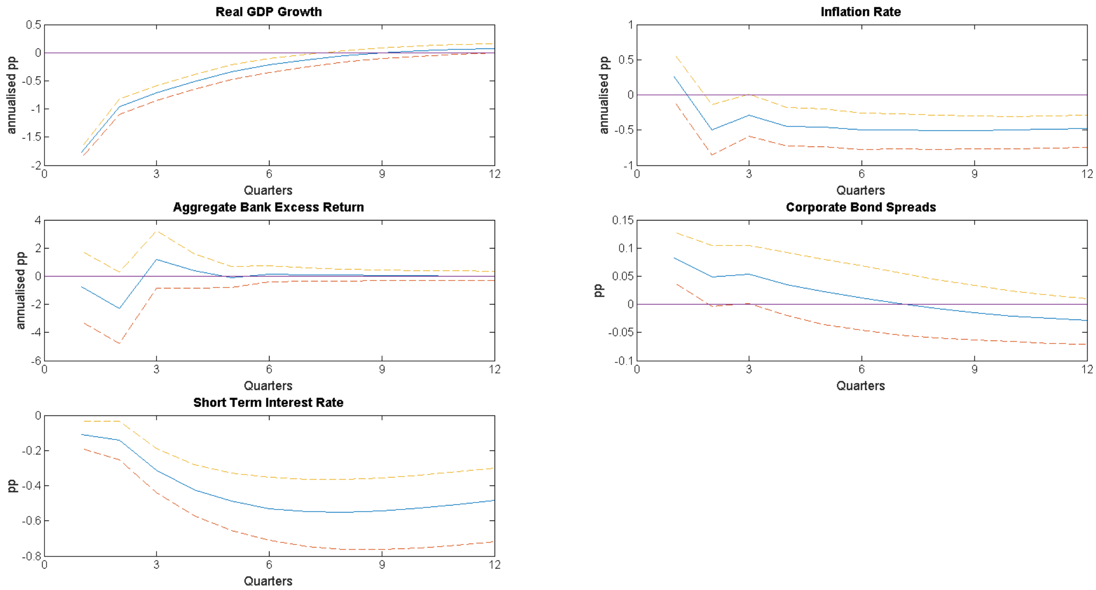
Notes: The error bands correspond to the 68% error bands. The red shaded area corresponds to the error bands of financial stress regimes. The grey shades correspond to those of financially non-stressful regimes. The green line is the median response of the BVAR model to the same shock.

Figure 3.6. Generalised Impulse Responses to a 1 SD adverse shock to corporate bond spreads in TVAR-S model



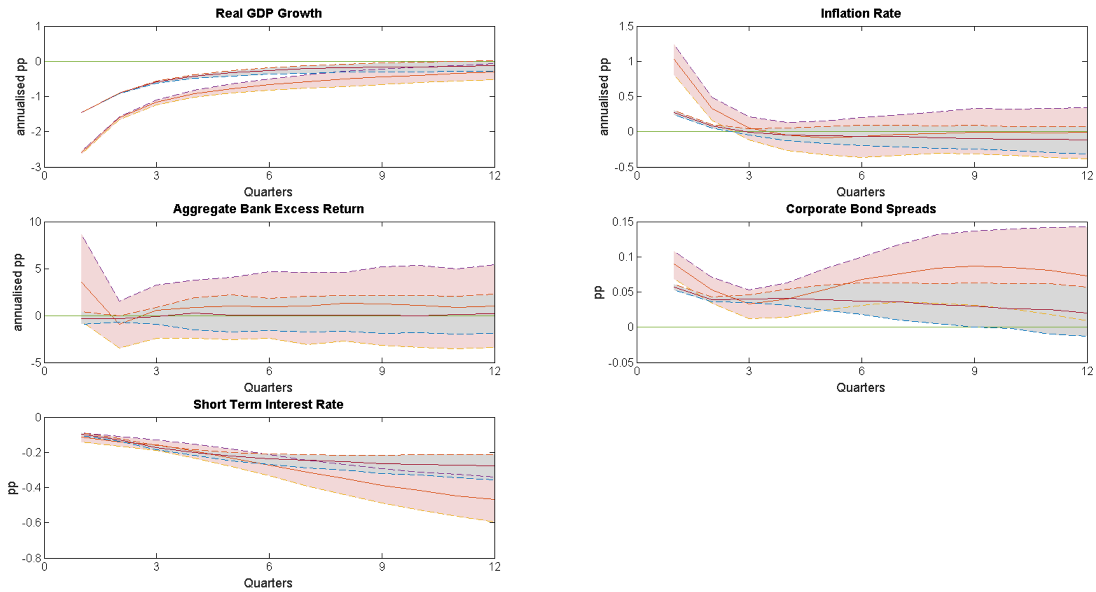
Notes: The error bands correspond to the 68% error bands. The red shaded area corresponds to the error bands of high stress regimes. The grey shades correspond to those of low stress regimes. The green line is the median response of the BVAR model to the same shock.

Figure 3.7. Generalised Impulse Responses to a 1 SD adverse shock to corporate bond spreads in MSVAR model



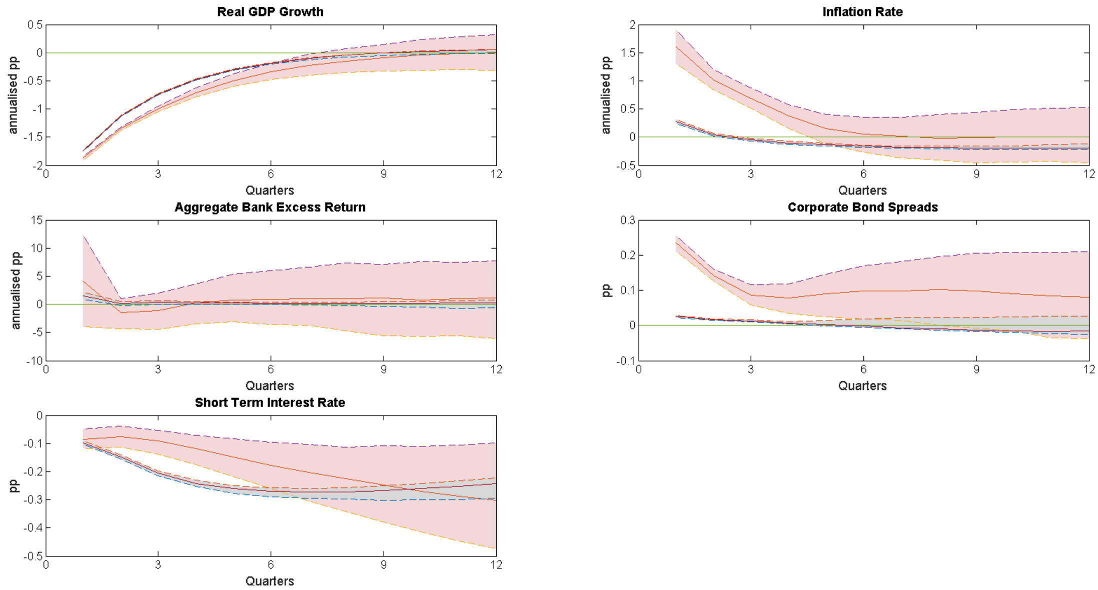
Notes: The error bands correspond to the 68% error bands.

Figure 3.8. Impulse Responses to a 1 SD adverse shock to real GDP growth in BVAR model



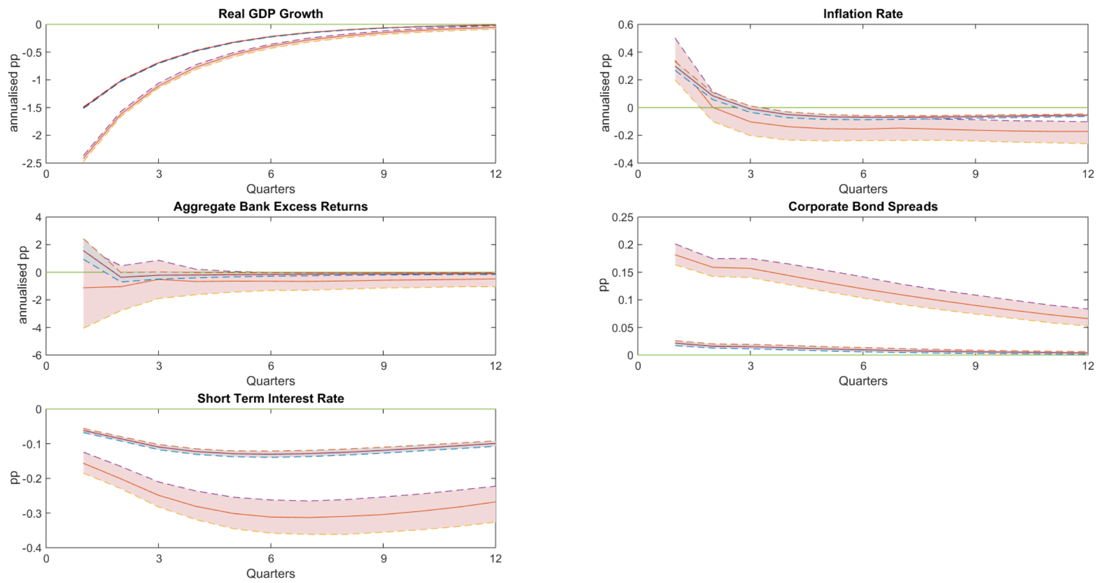
Notes: The error bands correspond to the 68% error bands. The red shaded area corresponds to the error bands of recessionary regimes. The grey shades correspond to those of non-recessionary regimes. The green line is the median response of the BVAR model to the same shock.

Figure 3.9. Generalised Impulse Responses to a 1 SD adverse shock to real GDP growth in TVAR-Y model



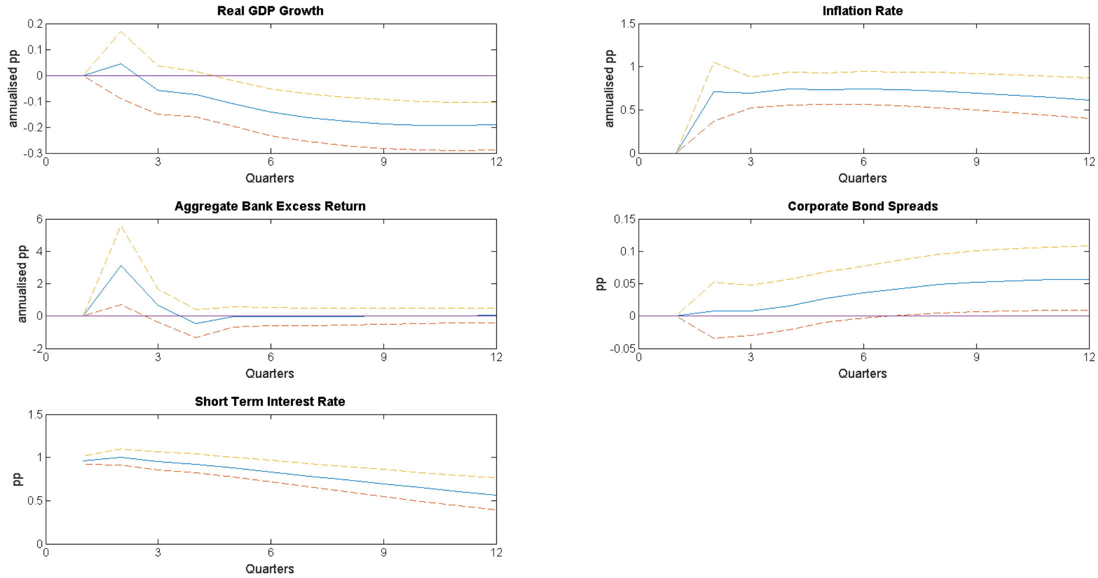
Notes: The error bands correspond to the 68% error bands. The red shaded area corresponds to the error bands of financial stress regimes. The grey shades correspond to those of financially non-stressful regimes. The green line is the median response of the BVAR model to the same shock.

Figure 3.10. Generalised Impulse Responses to a 1 SD adverse shock to real GDP growth in TVAR-S model



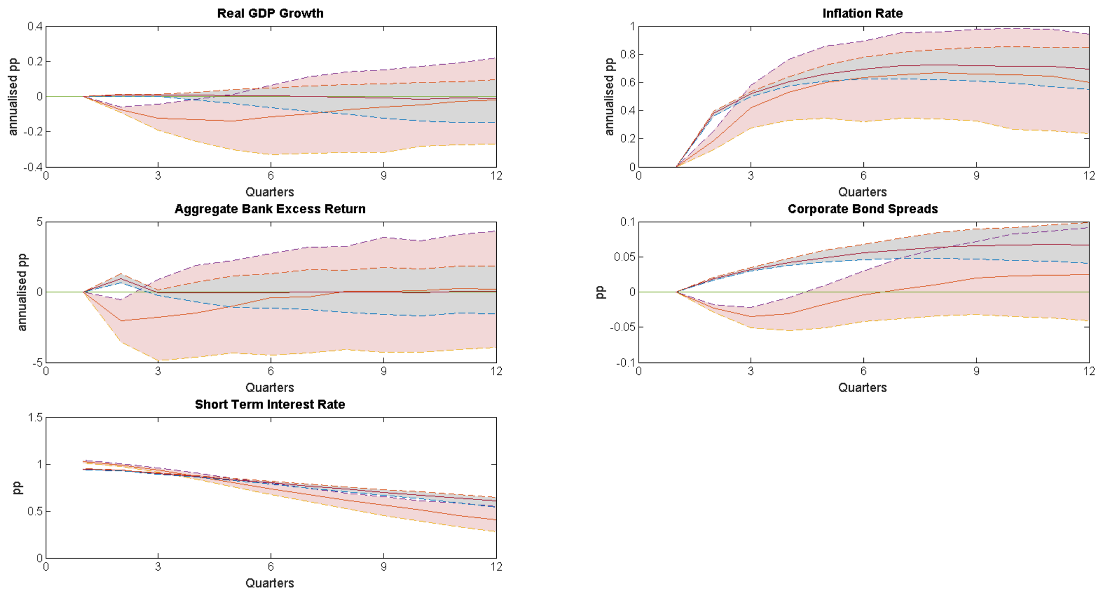
Notes: The error bands correspond to the 68% error bands. The red shaded area corresponds to the error bands of high stress regimes. The grey shades correspond to those of low stress regimes. The green line is the median response of the BVAR model to the same shock.

Figure 3.11. Generalised Impulse Responses to a 1 SD adverse shock to real GDP growth in MSVAR model



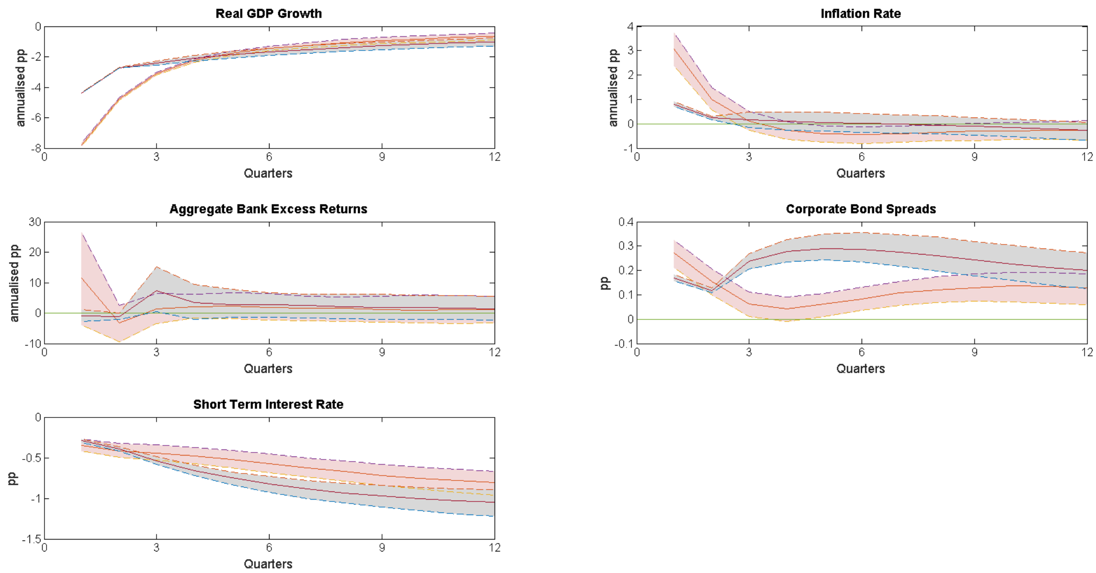
Notes: The error bands correspond to the 68% error bands.

Figure 3.12. Impulse Responses to a 1 SD adverse shock to short term interest rate in BVAR model



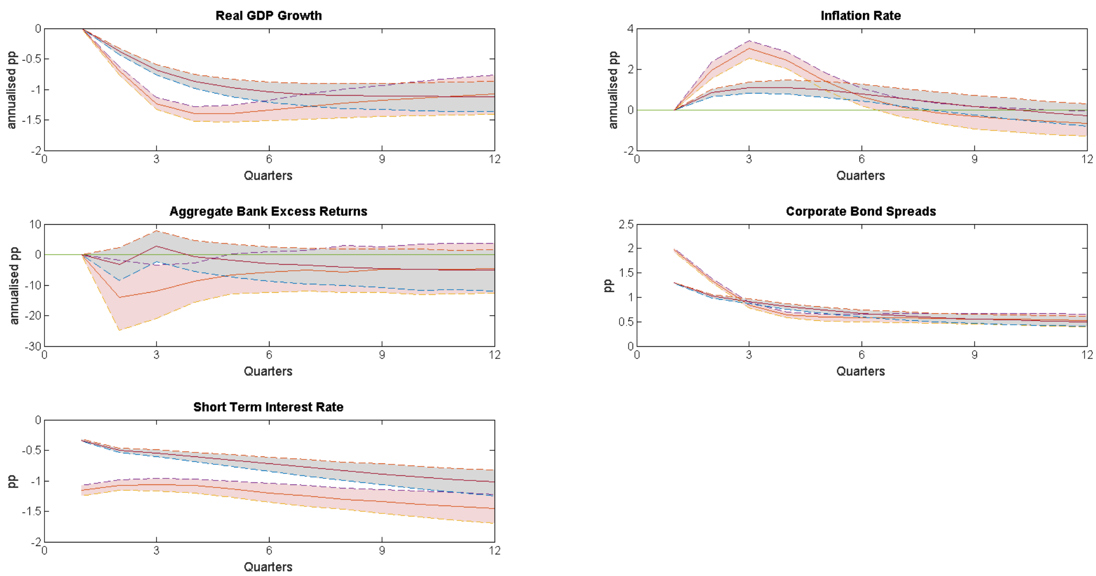
Notes: The error bands correspond to the 68% error bands. The red shaded area corresponds to the error bands of recessionary regimes. The grey shades correspond to those of non-recessionary regimes. The green line is the median response of the BVAR model to the same shock.

Figure 3.13. Generalised Impulse Responses to a 1 SD adverse shock to short term interest rate in TVAR-Y model



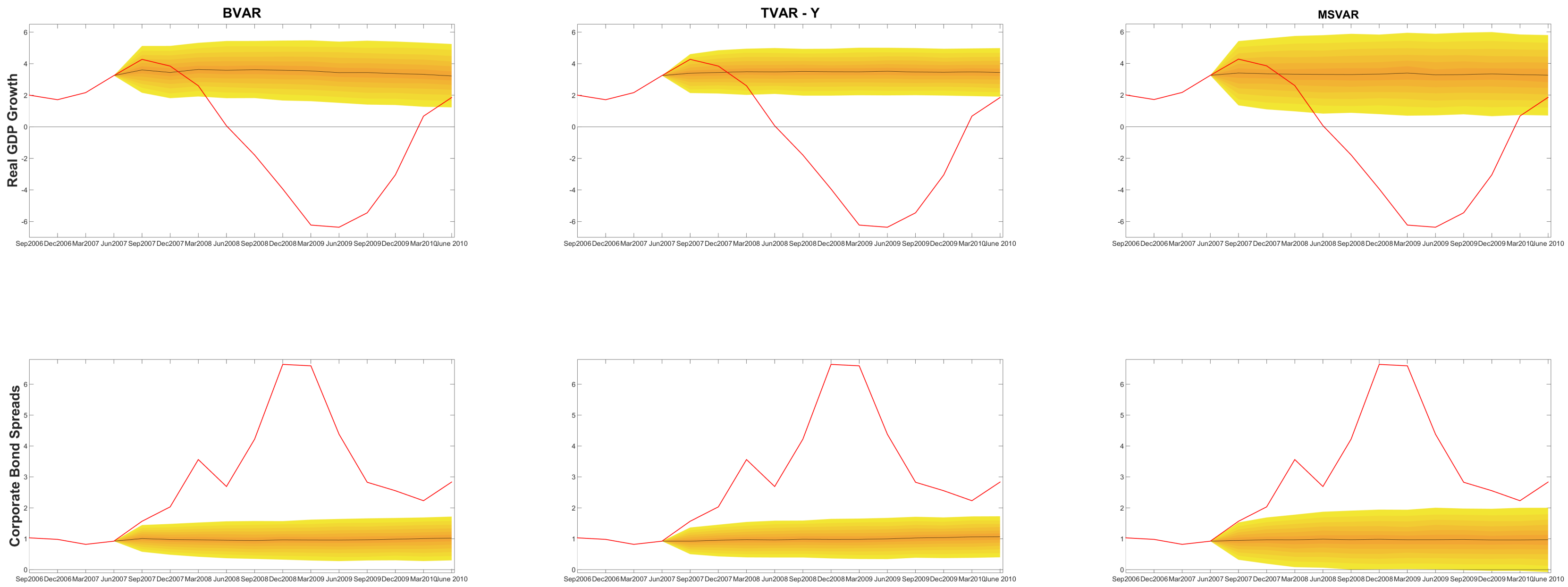
Notes: The error bands correspond to the 68% confidence intervals. The red shaded area corresponds to the error bands of recessionary regimes. The grey shades correspond to those of non-recessionary regimes.

Figure 3.14. Generalised Impulse Responses to a **3 SD adverse shock to real GDP growth** in TVAR-Y model (to be compared with Figure 3.9)



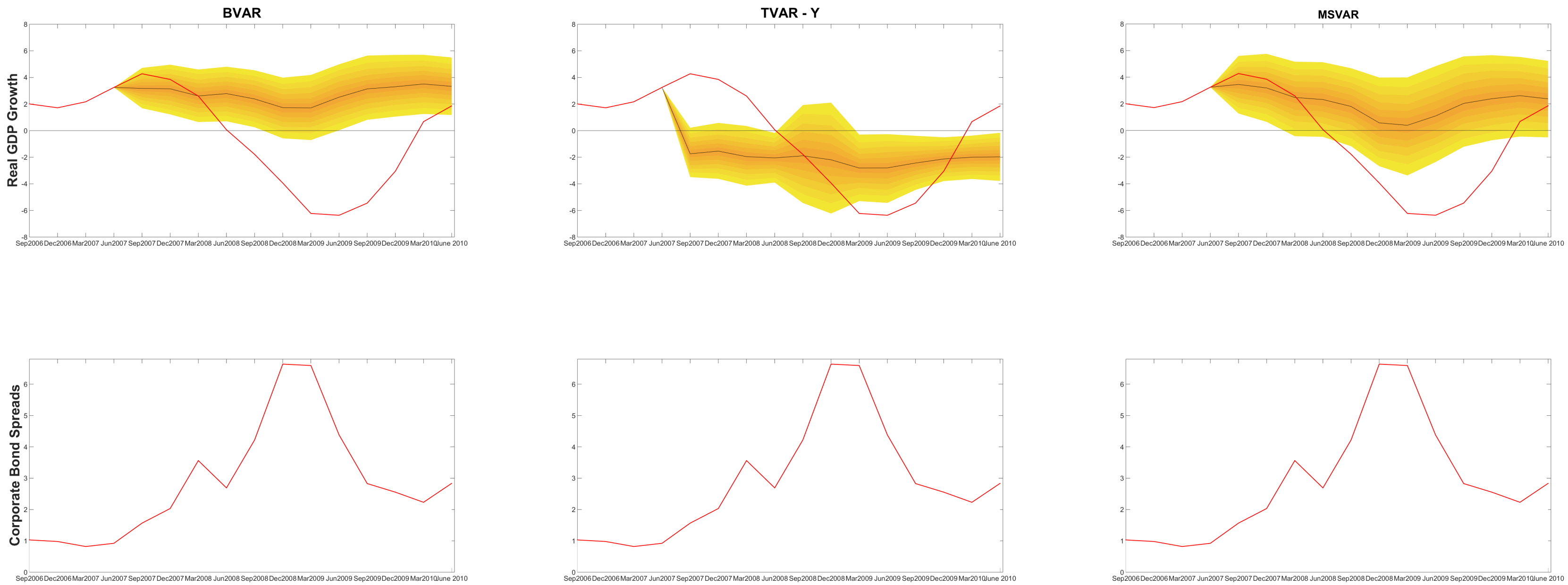
Notes: The error bands correspond to the 68% confidence intervals. The red shaded area corresponds to the error bands of financial stress regimes. The grey shades correspond to those of financially non-stressful regimes.

Figure 3.15. Generalised Impulse Responses to a **3 SD adverse shock to corporate bond spreads** in TVAR-S model (to be compared with Figure 3.6)



Notes: Unconditional forecast densities of BVAR, TVAR-Y, and MSVAR models are respectively shown in the the first, second and third columns. Actual out turns are indicated by the red lines. Fan charts indicate forecast bands between 20^{th} and 80^{th} percentile with 5% increments. The densities correspond to the 12-horizon pseudo-out-of-sample forecasts between 2007:Q3 and 2010:Q2, generated based on data between 1965:Q2 and 2007:Q2.

Figure 3.16. Unconditional predictive densities for the three models



Notes: Conditional forecast densities of BVAR, TVAR-Y, and MSVAR models are respectively shown in the the first, second and third columns. Actual out turns are indicated by the red lines. Fan charts indicate forecast bands between 20^{th} and 80^{th} percentile with 5% increments. The densities correspond to the 12-horizon pseudo-out-of-sample forecasts between 2007:Q3 and 2010:Q2, generated based on data between 1965:Q2 and 2007:Q2.

Figure 3.17. Conditional predictive densities based on the path of Corporate Bond Spreads

Appendix 3

3.A Normal Inverse Wishart priors

We impose normal inverse Wishart priors following Bańbura et al. (2010). For the prior means, we assume that the variables included in the VAR system follow an AR(1) process. The priors of the variance, on the other hand, has a sophisticated structure and defined as

$$\begin{aligned} \left(\frac{\lambda_1}{\ell\lambda_3}\right)^2 \quad & \text{if} \quad i = j \\ \left(\frac{\sigma_i \lambda_1 \lambda_2}{\sigma_j \ell^{\lambda_3}}\right)^2 \quad & \text{if} \quad i \neq j \\ (\sigma_i \lambda_4)^2 \quad & \text{for the constant} \end{aligned}$$

where i is the dependent variable in i^{th} equation and j is the independent variables in that equation. Therefore when $i = j$, it refers to the coefficients on the own lags of variable i. the variances σ_i and σ_j are the OLS estimations of the variances from AR regressions by using the VAR variables. The lag length in that particular step is shown by ℓ. The parameters λs are to control the tightness of the prior.

- λ_1 controls the standard deviation of the prior on own lags.
- λ_2 is the weight of own lag of dependent variable versus other lags. A value of unity implies that there is no difference on the lags of the dependent variable and other variables. It controls the standard deviation of the prior on lags of variables other than the dependent variable.
- λ_3 represents the lag decay. When its value increases, the coefficients on higher lags shrink to zero more tightly.
- λ_4 controls the prior variance on the constant term.

We choose our hyperparameters as $\lambda_1 = 0.1, \lambda_2 = 1, \lambda_3 = 1$ and $\lambda_4 = 10^5$ which are broadly similar to Canova (2007) and Blake and Mumtaz (2012).

3.B The Gibbs Sampling algorithm for TVAR

The following describes the Gibbs sampler for TVAR:

1. Given a value for the threshold variable, observations are separated into two regimes.

2. Given the observations in each regime, draw the coefficients and covariances.

3. Given values for coefficients and covariances, draw the threshold value.

$$Z^*_{new} = Z^*_{old} + \Psi^{1/2}\varepsilon$$

where $\Psi^{1/2}$ is abscaling factor and ε is distributed as $N(0, 1)$. Since the posterior distribution of the threshold value is not analytically tractable, we perform a Metropolis Hastings step, along with the Gibbs sampler. The scaling factor is chosen to ensure that the acceptance rate is in 20–40% interval.

4. Conditional on the threshold value, we sample the delay parameter d. Chen and Lee (1995) showed that the conditional posterior density of this parameter is multinomial distribution with probability $\frac{L(Y_t)}{\sum^d L(Y_t)}$ where $L(.)$ is the likelihood function. Note that we skip this step and fix d to our desired values based on our definitions of recessionary regimes and financially stressful regimes.

5. We run 100,000 draws and burn in the first 60,000 to ensure convergence.

3.C The Gibbs Sampling algorithm for MSVAR

The following describes the Gibbs sampling procedure:

1. Sampling the states, S_t:
Given values of VAR parameters and the covariances, we use multi–move Gibbs sampling proposed by Kim and Nelson (1999b). This method, conditional on data and the parameters, predicts the unobserved state and then updates it by running a simulation smoother in order to obtain a draw from the joint posterior densities, namely $f(S_t|Y_t, c_S, B_{1,S}, \ldots, B_{L,S}, P)$.

- We first calculate $f(S_T|Y_T)$. Hamilton (1989) provides a filter to evaluate $f(S_t|Y_t)$ for $t = 1, 2, \ldots, T$.
- We then calculate $f(S_t|S_{t+1}, Y_t)$. Kim and Nelson (1999b) show that

$$f(S_t|S_{t+1}, Y_t) \propto f(S_t|S_{t+1})f(S_t|Y_t) \tag{3.3}$$

  where the Hamilton filter provides $f(S_t|Y_t)$ and $f(S_t|S_{t+1})$.

2. Sampling the covariances, Ω_S:

   Given the states, we sample the covariance matrices from inverse Wishart distribution,

$$\Omega_S \sim iW(\bar{H}_S, \varphi_S)$$

   where $S = 1, 2$. $\bar{H}_S$ refers to the covariance matrix in regime S, and the parameter φ_S refers to the number of the observations in each regime.

3. Sampling the VAR coefficients, $c_{S_t}, B_{1,S_t}, B_{2,S_t}, \ldots, B_{L,S_t}$:

   Given the states and the covariances, we sample $c_{S_t}, B_{1,S_t}, B_{2,S_t}, \ldots, B_{L,S_t}$.

   The conditional posterior of the VAR coefficients is

$$vec(B) \mid \Omega_s, Y \sim N(vec(\tilde{B}), \Omega_S \otimes (X^{*'}X^*)^{-1}).$$

4. Sampling the transition probabilities, P:

   As the last step we sample the transition probabilities. We impose Dirichlet priors for the non–zero elements of the transition matrix, p_{ij}:

$$p^0_{ij} = D(u_{ij})$$

   where $D(.)$ represents the Dirichlet distribution. The posterior distributions of the transition probabilities are

$$p_{ij} = D(u_{ij} + \eta_{ij})$$

   where η_{ij} denotes the number of times regime i is followed by regime j. The value of u_{ij} equals 20, which implies a prior belief that probability of staying in the same regime to be is 0.85.

5. We employ 50,000 iterations for the Gibbs sampling. We discard the first 10,000 draws as burn in and keep every tenth draw to ensure convergence.

3.C.1 An extension: MSVAR with two independent Markov chains

We can extend the MSVAR system by exploiting two independent Markov chains separately for variance and slope coefficients, as in Barnett et al. (2010). Again, we assume two regimes for each chains.

In this specific case, we can also make use of time varying transition probabilities which are indicated by the subscript t. Naturally, the representation of the model and the Gibbs sampling algorithm have to be modified accordingly. Here, the MSVAR model is written as

$$Y_t = c_{S_t} + B_{1,S_t} Y_{t-1} + B_{2,S_t} Y_{t-2} + \cdots + B_{l,S_t} Y_{t-L} + v_t, \tag{3.4}$$

where $e_t \sim N(0, \Omega_{s_t})$. This VAR model incorporates regime changes both in its coefficients, denoted by $S_t = 1, ..., M$, and the variance of the error terms, denoted by $s_t = 1, ..., m$. The changes of coefficient and variance regimes are independent of each other.

The first set of latent regimes S_t are denoted as $S = 1, 2$ which we refer as variance regimes. The regimes associated with the coefficients are denoted as $s = 3, 4$. They are assumed to follow two independent first order Markov chains. Therefore we have two transition matrices, one for each regime, P and Q,

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \text{ and } \begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix}$$

where $p_{ij} = prob(S_t = i | S_{t-1} = j)$ indicates regime i is followed by regime j in variance regimes and $q_{ij} = prob(s_t = i | s_{t-1} = j)$ indicates regime i is followed by regime j in coefficient regimes. The probability of high variance regime is followed by low variance regime is p_{21} while q_{21} corresponds to high coefficient regime being followed by low coefficient regime. The columns sum up to 1.

The Gibbs sampling is more involved because we have to take into account of independent switches of regimes for variances and slope coefficients. Conditional on the variance regimes and the corresponding covariances, we rewrite the model as:

$$Y^* = A_{j,s}X^* + V_t \tag{3.5}$$

where $Y^* = I(S_t = 1)[(\Omega_1^{-1} \otimes I_t^*)^{-1/2} \times vec(Y_t)] + I(S_t = 2)[(\Omega_2^{-1} \otimes I_t^*)^{-1/2} \times vec(Y_t)]$, $X^* = I(S_t = 1)[(\Omega_1^{-1} \otimes I_t^*)^{-1/2} \times (X_t \otimes I_N)] + I(S_t = 2)[(\Omega_2^{-1} \otimes I_t^*)^{-1/2} \times (X_t \otimes I_N)]$ where Y_t includes the observations corresponding to the relevant variance regimes. This new representation is an MSVAR model with homoscedastic covariance matrix. We can make use of multi–move Gibbs sampling to draw the coefficient states, $f(s_t|Y_t, c_S, B_{1,S}, \ldots, B_{L,S}, P, Q)$.

Similarly, we sample the variance regimes and covariance matrices conditional on the slope coefficient regimes and the sampled coefficients.

The estimation results of this extension is given by Figure 3.1. The first set of regimes are identified as high variance regimes given that they successfully capture high volatility periods before Great Moderation and around the Great Recession. The second set of the regimes are attributed to the high and low mean states. High mean states appear to capture the high inflation and interest rates periods before the 1990s.

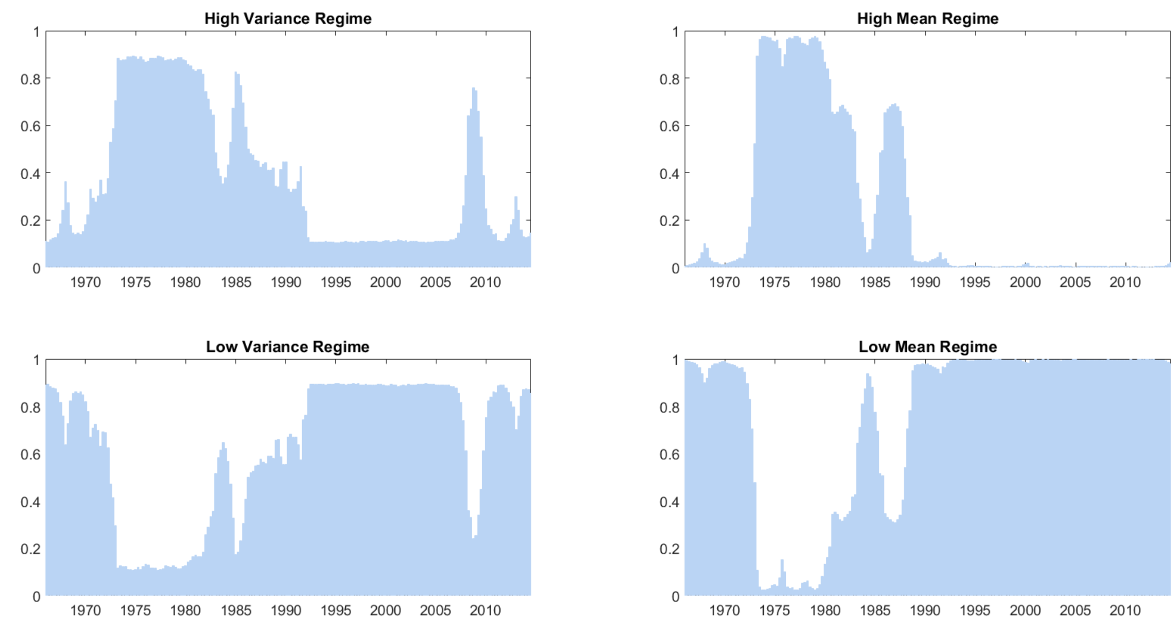


Figure 3.1. Full sample regimes for MSVAR with 2 independent Markov chains

3.D Generalized Impulse Response Functions

We compute the nonlinear impulse response functions of MSVAR and TVAR models by following Koop et al. (1996), Baum and Koester (2011) and Afonso et al. (2011).

3.D.1 GIRFs for TVAR

The following steps are separately employed for each regime for both TVAR–Y and TVAR–S models .

1. Run the estimation and save all parameter draws.
2. Given a Gibbs draw, pick a random history from the set recessionary/financially stressful observations.
3. Draw random shocks and form a set of unconditional forecasts which are denoted as y_{t+k}^{Th} where Th indicates the TVAR model and k is forecast horizon. The output is a $(horizon \times N)$ matrix of forecasts for all N variables and these forecasts serve as a baseline.
4. Form another set of forecasts with the same random shocks except that a specific shock is perturbed at horizon 0. Refer these forecasts as $y_{t+k}^{Th,p}$. The output is a $(horizon \times N \times 1)$ matrix *for a given shock*. If one is interested in shocking all the variables, the resulting matrix is size of $(horizon \times N \times N)$.
5. Repeat steps 3 to 5 for $Simm = 500$.
6. Take the means of the forecasts over $Simm$ and calculate the difference between the means such that $\frac{1}{Simm} \sum_{Simm} y_{t+k}^{Th,p} - \frac{1}{Simm} \sum_{Simm} y_{t+k}^{Th}$.
7. Repeat steps 3 to 7 for all Gibbs draws and all histories. The result of this step is the time varying impulse response functions.
8. Take the mean of the resulting impulse response functions from all Gibbs draws. The output is the ultimate GIRFs of recessionary regime in TVAR model.
9. Repeat steps 3 to 9 for the non-recessionary/financially stressful regimes.

3.D.2 GIRFs for MSVAR

1. Run the estimation and save all parameter draws.

2. Given a Gibbs draw and at time t, projecting the ergodic probabilities and draw random shocks to form a set of unconditional forecasts y^{M}_{t+k} where k is the forecast horizon and superscript M marks the MSVAR model. The output is a $(horizon \times N)$ matrix of forecasts for all N variables and these forecasts serve as a baseline.

3. Form another set of forecasts with the same random shocks except with perturbed shocks at horizon 0. Refer these forecasts as $y^{M,p}_{t+k}$ where the additional superscript addresses the perturbed shocks. The output is a $(horizon \times N \times 1)$ matrix *for a given shock*. If one is interested in shocking all the variables, the resulting matrix is size of $(horizon \times N \times N)$.

4. Repeat both steps 2 and 3 for $Simm = 500$.

5. Take the mean of the resulting forecasts over Simm and the difference between the means such that $\frac{1}{Simm}\sum_{Simm} y^{M,p}_{t+k} - \frac{1}{Simm}\sum_{Simm} y^{M}_{t+k}$. This difference is for a given Gibbs draw and given time period.

6. Repeat steps 2 to 5 for all Gibbs draws and for all $t = 1, 2, 3, ..., T$.

7. Take the mean of the time varying impulse response functions from the previous step over the high stress regimes as identified by the model. The output of this step gives the GIRFs of MSVAR model.

3.E Variables

We us *Global Financial Data* to construct the following variables: real GDP growth rate (mnemonic: GDPCGBR), inflation rate (mnemonic: CPGBRCM) and the short-term interest rate (mnemonic: ITGBR3D). To construct the aggregate credit bond spread series, we take the difference between UK corporate bond yield (mnemonic: INGBRW) and UK 10 year government bond yield (mnemonic: IGGBR10D).

Data for bank excess returns, UK equity index (mnemonic: TOTMKUK) and UK banks equity index (mnemonic: BANKSUK), are taken from *DataStream*.

The tables in this section give the descriptive statistics and the correlation matrix of these variables. The charts for these variables for the whole data span of 1965:Q2 to 2014:Q2 are also given below. The dashed vertical lines indicate the observation in 2007:Q2 which is the quarter when we separate the data for *conditional* forecasting purposes.

Table 3.1. Summary statistics

	Real GDP Growth	Inflation Rate	Agg. Bank Excess Returns	Corporate Bond Spreads	Short Term Interest Rate
Mean	2.39	5.55	1.26	1.30	7.19
Median	2.60	3.88	0.43	1.03	6.56
Maximum	12.78	41.90	120.85	6.64	16.27
Minimum	-6.52	-5.96	-112.36	-0.33	0.32
Std Deviation	2.49	6.11	33.34	1.02	3.88
Skewness	-0.82	2.06	0.11	2.07	0.15
Kurtosis	6.23	9.78	4.53	9.67	2.54
Observations	197	197	197	197	197

Table 3.2. Contemporaneous correlation coefficients of variables

	Real GDP Growth	Inflation Rate	Agg. Bank Excess Returns	Corporate Bond Spreads	Short Term Interest Rate
Real GDP Growth	1				
Inflation Rate	-0.2070	1			
Agg. Bank Excess Returns	0.0053	0.0223	1		
Corporate Bond Spreads	-0.4356	0.2302	-0.0151	1	
Short Term Interest Rate	0.0002	0.4850	0.0001	-0.1141	1

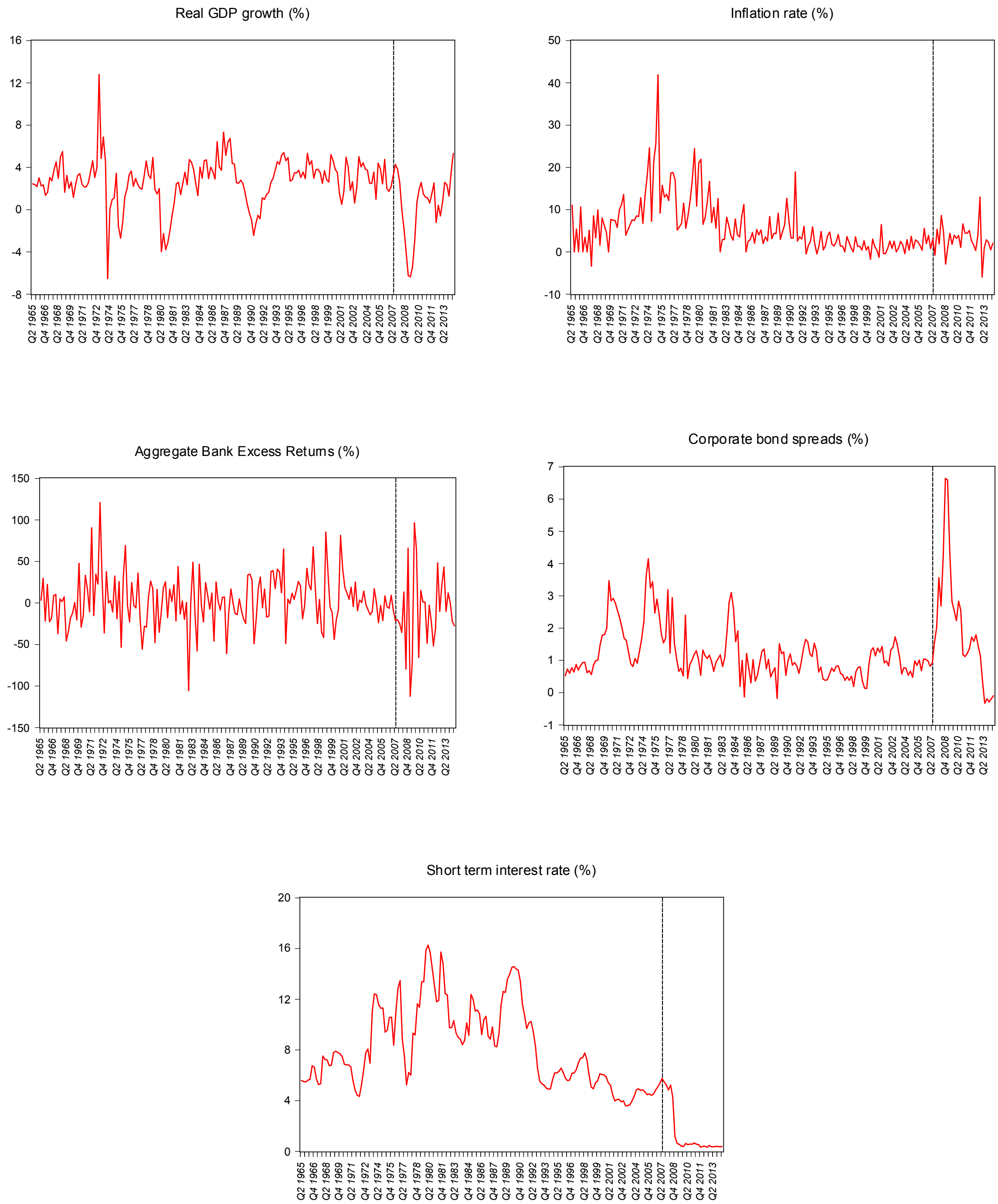


Figure 3.2. Variables

Bibliography

Adrian, T. and M. K. Brunnermeier (2014). Covar. Technical report, Federal Reserve Bank of New York Staff Report 348.

Afonso, A., J. Baxa, and M. SlavÃk (2011, March). Fiscal developments and financial stress: a threshold VAR analysis. Working Papers Department of Economics 2011/11, ISEG - School of Economics and Management, Department of Economics, University of Lisbon.

Aguilar, O. and M. West (2000). Bayesian dynamic factor models and portfolio allocation. *Journal of Business and Economic Statistics 18*, 338–357.

Ahmadi, P. A. (2009). Financial Shocks, Monetary Policy, and Business Cycles: Evidence from a Structural Time Varying Bayesian FAVAR. Job market paper, Humbolt University of Berlin.

Aikman, D., P. Alessandri, B. Eklund, P. Gai, S. Kapadia, E. Martin, N. Mora, G. Sterne, and M. Willison (2011, August). Funding Liquidity Risk in a Quantitative Model of Systemic Stability. In R. Alfaro (Ed.), *Financial Stability, Monetary Policy, and Central Banking*, Volume 15 of *Central Banking, Analysis, and Economic Policies Book Series*, Chapter 12, pp. 371–410. Central Bank of Chile.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In B. Petrow and F. Csake (Eds.), *Second International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado.

Alessandri, P. and H. Mumtaz (2014). Financial conditions and density forecasts for us output and inflation. Technical report, Working Paper, School of Economics and Finance, Queen Mary, University of London.

Alessi, L., M. Barigozzi, and M. Capasso (2010). Improved penalization for determining the number of factors in approximate factor models. *Statistics & Probability Letters 80*(23), 1806–1813.

Alexander, M. and N. Christakis (2008). Bias and asymmetric loss in expert forecasts: A study of physician prognostic behavior with respect to patient survival. *Journal of health economics 27*(4), 1095–1108.

Andersen, P. and R. Gill (1982). Cox's regression model for counting processes: A large sample study. *Annals of Statistics 10*(4), 1100–1120.

Artis, M. and M. Marcellino (2001). Fiscal forecasting: The track record of the IMF, OECD and EC. *The Econometrics Journal 4*(1), 20–36.

Artis, M. J., A. Banerjee, and M. Marcellino (2005). Factor forecasts for the UK. *Journal of Forecasting 24*(4), 279–298.

Auffhammer, M. (2007). The rationality of EIA forecasts under symmetric and asymmetric loss. *Resource and Energy Economics 29*(2), 102–121.

Ayebo, A. and T. J. Kozubowski (2003). An asymmetric generalization of Gaussian and Laplace laws. *Journal of Probability and Statistical Science 1*(2), 187–210.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics 12*(2), 171–178.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica 71*(1), 135–171.

Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. *Econometrica 70*(1), 191–221.

Bai, J. and S. Ng (2006). Confidence intervals for diffusion index forecasts and inference for factor-augmented regressions. *Econometrica 74*(4), 1133–1150.

Bai, J. and S. Ng (2008a). Forecasting economic time series using targeted predictors. *Journal of Econometrics 146*(2), 304–317.

Bai, J. and S. Ng (2008b). Large dimensional factor analysis. *Foundations and Trends® in Econometrics 3*(2), 89–163.

Bai, J. and S. Ng (2011). Principal components estimation and identification of the factors. Working Paper 1112-02, Department of Economics, Columbia University.

Bai, J. and P. Wang (2011, August). Conditional Markov chain and its application in economic time series analysis. MPRA Paper 33369, University Library of Munich, Germany.

Balke, N. S. and K. M. Emery (1994). Understanding the price puzzle. *Economic and Financial Policy Review* (Q IV), 15–26.

Bańbura, M., D. Giannone, and M. Lenza (2015a). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting*.

Bańbura, M., D. Giannone, and M. Lenza (2015b). Conditional forecasts and scenario analysis with vector autoregressions for large cross-sections. *International Journal of Forecasting*.

Bańbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. *Journal of Applied Econometrics 25*(1), 71–92.

Banerjee, A., M. Marcellino, and I. Masten (2008). Forecasting macroeconomic variables using diffusion indexes in short samples with structural change. Working Papers 334, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University.

Bank of England (2013). A framework for stress testing the UK banking system. a discussion paper. Technical report.

Bank of England (2014). Stress testing the uk banking system: key elements of the 2014 stress test. Technical report.

Bank of England (2015). The bank of england's approach to stress testing the UK banking system. Technical report.

Bank of England (2015). *The Bank of England's approach to stress testing the UK banking system*. Bank of England.

Bank of England (2015). Stress testing the uk banking system: key elements of the 2015 stress test. Technical report.

Barnett, A., J. J. Groen, and H. Mumtaz (2010). Time-varying inflation expectations and economic fluctuations in the United Kingdom: a structural VAR analysis.

Barnett, A., H. Mumtaz, and K. Theodoridis (2012, May). Forecasting UK GDP growth, inflation and interest rates under structural change: a comparison of models with time-varying parameters. Bank of England working papers 450, Bank of England.

Basu, S. and S. Markov (2004a). Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics 38*, 171–203.

Basu, S. and S. Markov (2004b). Loss function assumptions in rational expectations tests on financial analysts' earnings forecasts. *Journal of Accounting and Economics 38*, 171–203.

Batchelor, R. and D. Peel (1998). Rationality testing under asymmetric loss. *Economics Letters 61*(1), 49–54.

Baum, A. and G. B. Koester (2011). The impact of fiscal policy on economic activity over the business cycle - evidence from a threshold VAR analysis. Discussion Paper Series 1: Economic Studies 2011,03, Deutsche Bundesbank, Research Centre.

Baumeister, C., P. Liu, and H. Mumtaz (2010, October). Changes in the transmission of monetary policy: evidence from a time-varying factor-augmented VAR. Bank of England working papers 401, Bank of England.

Belloni, A. and V. Chernozhukov (2013). Least squares after model selection in high-dimensional sparse models. *Bernoulli 19*(2), 521–547.

Belviso, F. and F. Milani (2006). Structural factor-augmented VARs (SFAVARs) and the effects of monetary policy. *The B.E. Journal of Macroeconomics 6*, 1–46.

Bernanke, B., J. Boivin, and P. S. Eliasz (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics 120*, 387– 422.

Bernanke, B. S. and A. S. Blinder (1992, September). The federal funds rate and the channels of monetary transmission. *American Economic Review 82*(4), 901–21.

Bernanke, B. S., M. Gertler, and S. Gilchrist (1999). Chapter 21 the financial accelerator in a quantitative business cycle framework. Volume 1, Part C of *Handbook of Macroeconomics*, pp. 1341 – 1393. Elsevier.

Blake, A. and H. Mumtaz (2012). *Technical Handbook – No. 4 Applied Bayesian econometrics for central bankers*. Threadneedle Street, London, EC2R 9AH: Centre for Central Banking Studies, Bank of England.

Board of Governers of the Federal Reserve System (2015). Dodd-frank act stress test 2015: Supervisory stress test methodology and results. Technical report.

Boero, G., J. Smith, and K. Wallis (2008). Evaluating a three-dimensional panel of point forecasts: the bank of england survey of external forecasters. *International Journal of Forecasting 24*(3), 354–367.

Boissay, F., F. Collard, and F. Smets (2013). Booms and systemic bank crises. Technical report, European Central Bank working paper 1514.

Boivin, J. and S. Ng (2006). Are more data always better for factor analysis? *Journal of Econometrics 132*, 169–194.

Bonham, C. and R. Cohen (1995). Testing the rationality of price forecasts: Comment. *The American Economic Review 85*(1), 284–289.

Bork, L. (2009). Estimating US Monetary Policy Shocks Using a Factor Augmented Vector Autoregression: An EM Approach. Research paper, Aarhus University, School of Economics and Management.

Boucher, C. and B. Maillet (2015). La macroÃ©conomie-en-risque. *Revue Ã©conomique 66*(4), 769–782.

Breuer, T., M. Jandacka, K. Rheinberger, and M. Summer (2009, September). How to Find Plausible, Severe and Useful Stress Scenarios. *International Journal of Central Banking 5*(3), 205–224.

Brunnermeier, M. and Y. Sannikov (2014). A macroeconomic model with a financial sector. *American Economic Review 104(2)*, 379–421.

Cain, M. and C. Janssen (1995). Real estate price prediction under asymmetric loss. *Annals of the Institute of Statistical Mathematics 47*(3), 401–414.

Calabria, R. and G. Pulcini (1996). Point estimation under asymmetric loss functions for left-truncated exponential samples. *Communications in Statistics-Theory and Methods 25*(3), 585–600.

Canova, F. (2007). *Methods for applied macroeconomic research*, Volume 13. Princeton University Press.

Capistrán, C. (2008). Bias in Federal Reserve inflation forecasts: Is the Federal Reserve irrational or just cautious? *Journal of Monetary Economics 55*(8), 1415–1427.

Carter, C. K. and R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika 81*(3), 541–55.

Castelnuovo, E. and P. Surico (2010, December). Monetary Policy, Inflation Expectations and The Price Puzzle. *Economic Journal 120*(549), 1262–1283.

Chen, C. W. and J. C. Lee (1995). Bayesian inference of threshold autoregressive models. *Journal of Time Series Analysis 16*(5), 483–492.

Christiano, L. J., M. Eichenbaum, and C. Evans (1996, February). The Effects of Monetary Policy Shocks: Evidence from the Flow of Funds. *The Review of Economics and Statistics 78*(1), 16–34.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1998, February). Monetary policy shocks: What have we learned and to what end? Working Paper 6400, National Bureau of Economic Research.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (2005, February). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy 113*(1), 1–45.

Christodoulakis, G. (2005). Financial forecasts in the presence of asymmetric loss aversion, skewness and excess kurtosis. *Finance Research Letters 2*(4), 227–233.

Christodoulakis, G. and E. Mamatzakis (2008). An assessment of the EU growth forecasts under asymmetric preferences. *Journal of Forecasting 27*(6), 483–492.

Christodoulakis, G. and E. Mamatzakis (2009). Assessing the prudence of economic forecasts in the EU. *Journal of Applied Econometrics 24*(4), 583–606.

Christoffersen, P. and F. Diebold (1996). Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics 11*(5), 561–571.

Christoffersen, P. F. and F. X. Diebold (1997). Optimal prediction under asymmetric loss. *Econometric Theory 13*(06), 808–817.

Clark, T. E. (2009). Is the Great Moderation over? An empirical analysis. *Economic Review 4*, 5–42.

Clatworthy, M., D. Peel, and P. Pope (2006). Are analysts' loss functions asymmetric? *Journal of Forecasting*.

Clements, M. (1997). Evaluating the rationality of fixed-event forecasts. *Journal of Forecasting 16*(4), 225–239.

Clements, M. and D. Hendry (1993). On the limitations of comparing mean square forecast errors. *Journal of Forecasting 12*(8), 617–637.

Clements, M., F. Joutz, and H. Stekler (2007). An evaluation of the forecasts of the Federal Reserve: A pooled approach. *Journal of Applied Econometrics 22*(1), 121–136.

Clements, M. and N. Taylor (2001). Robust evaluation of fixed-event forecast rationality. *Journal of Forecasting 20*(4), 285–295.

Clements, M. P. and J. Smith (2000). Evaluating the forecast densities of linear and non-linear models: applications to output growth and unemployment. *Journal of Forecasting 19*(4), 255–276.

Covas, F. B., B. Rump, and E. ZakrajÅ¡ek (2014). Stress-testing US bank holding companies: A dynamic panel quantile regression approach. *International Journal of Forecasting 30*(3), 691–713.

D'Agostino, A. and D. Giannone (2006). Comparing alternative predictors based on large-panel factor models. Working Paper Series 0680, European Central Bank.

De Nicolo, G. and M. Lucchetta (2016). Forecasting tail risks. *Journal of Applied Econometrics*, http://onlinelibrary.wiley.com/doi/10.1002/jae.2509/pdf.

Demetrescu, M. (2006). An extension of the Gauss–Newton algorithm for estimation under asymmetric loss. *Computational Statistics & Data Analysis 50*(2), 379–401.

den Reijer, A. H. J. (2005). Forecasting Dutch GDP using large scale factor models. DNB Working Papers 28, Netherlands Central Bank, Research Department.

Dias, F., M. Pinheiro, and A. Rua (2010). Forecasting using targeted diffusion indexes. *Journal of Forecasting 29*(3), 341–352.

Diebold, F. X. and R. S. Mariano (1995). Comparing predictive accuracy. *Journal of Business & Economic Statistics 13*(3), 253–263.

Drehmann, M., A. J. Patton, and S. Sorensen (2007). Non-linearities and stress testing. *Risk Measurement and Systemic Risk*.

Eichenbaum, M. (1992). Comments on interpreting the macroeconomic time series facts: the effects of monetary policy by c.a. sims. *European Economic Review 36*(5), 1001 – 1011.

Eickmeier, S., W. Lemke, and M. Marcellino (2011a). Classical time-varying FAVAR models - estimation, forecasting and structural analysis. Discussion Paper Series 1: Economic Studies 2011,04, Deutsche Bundesbank, Research Centre.

Eickmeier, S., W. Lemke, and M. Marcellino (2011b). The changing international transmission of financial shocks: evidence from a classical time-varying FAVAR. Discussion Paper Series 1: Economic Studies 2011,05, Deutsche Bundesbank, Research Centre.

Elliott, G., I. Komunjer, and A. Timmermann (2005a, Oct.). Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies 72*(4), 1107–1125.

Elliott, G., I. Komunjer, and A. Timmermann (2005b). Estimation and testing of forecast rationality under flexible loss. *Review of Economic Studies 72*(4), 1107–1125.

Elliott, G., I. Komunjer, and A. Timmermann (2008). Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association 6*(1), 122–157.

Elliott, G. and A. Timmermann (2004). Optimal forecast combinations under general loss functions and forecast error distributions. *Journal of Econometrics 122*(1), 47–79.

Elliott, G. and A. Timmermann (2008). Economic forecasting. *Journal of Economic Literature 46*(1), 3–56.

Engel, C., N. C. Mark, and K. D. West (2012). Factor model forecasts of exchange rates. Working Paper 18382, National Bureau of Economic Research.

Fernandez, C., J. Osiewalski, and M. F. Steel (1995). Modeling and inference with υ-spherical distributions. *Journal of the American Statistical Association 90*(432), 1331–1340.

Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model: One-sided estimation and forecasting. *Journal of the American Statistical Association 100*, 830–840.

Gilchrist, S. and E. Zakrajšek (2012). Credit spreads and business cycle fluctuations. *American Economic Review 102*(4), 1692–1720.

Gneiting, T. (2011). Quantiles as optimal point forecasts. *International Journal of Forecasting 27*(2), 197–207.

Godbout, C. and M. J. Lombardi (2012). Short-term forecasting of the Japanese economy using factor models. Working Paper Series 1428, European Central Bank.

Granger, C. (1999). Outline of forecast theory using generalized cost functions. *Spanish Economic Review 1*(2), 161–173.

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational Research Society 20*(2), 199–207.

Hacioglu Hoke, S. and K. Tuzcuoglu (2014). Interpreting the latent dynamic factors by threshold FAVAR model. University of Bonn Working Paper.

Hamilton, J. D. (1989, March). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica 57*(2), 357–84.

Hoggarth, G., S. Sorensen, and L. Zicchino (2005, November). Stress tests of UK banks using a VAR approach. Bank of England working papers 282, Bank of England.

Hubrich, K. and T. Terasvirta (2013, June). Thresholds and Smooth Transitions in Vector Autoregressive Models. CREATES Research Papers 2013-18, School of Economics and Management, University of Aarhus.

Hubrich, K. and R. J. Tetlow (2015). Financial stress and economic dynamics: The transmission of crises. *Journal of Monetary Economics 70*, 100–115.

Hürtgen, P. and J. Cloyne. The macroeconomic effects of monetary policy: A new measure for the United Kingdom. *American Economic Journal: Macroeconomics*. Forthcoming.

Kaabia, O. and I. Abid (2013, March/April). Theoretical Channels of International,Transmission During the Subprime Crisis to OECD Countries: A FAVAR Model Under Bayesian Framework. *The Journal of Applied Business Research 29*(2), 443–460.

Kadiyala, K. R. and S. Karlsson (1997, March-Apr). Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics 12*(2), 99–132.

Kalliovirta, L., M. Meitz, and P. Saikkonen (2014). Gaussian mixture vector autoregression.

Kapinos, P. and O. A. Mitnik (2015). A top-down approach to stress-testing banks. *Journal of Financial Services Research*, 1–36.

Keane, M. and D. Runkle (1990). Testing the rationality of price forecasts: New evidence from panel data. *The American Economic Review 80*, 714–735.

Kilian, L. and S. Manganelli (2008). The central banker as a risk manager: Estimating the federal reserve's preferences under greenspan. *Journal of Money, Credit and Banking 40*(6), 1103–1129.

Kim, C.-J. and C. R. Nelson (1999a, November). Has The U.S. Economy Become More Stable? A Bayesian Approach Based On A Markov-Switching Model Of The Business Cycle. *The Review of Economics and Statistics 81*(4), 608–616.

Kim, C.-J. and C. R. Nelson (1999b). *State-space models with regime switching.* Cambridge, Massachusetts: MIT Press.

Kiyotaki, N. and J. Moore (1997, April). Credit Cycles. *Journal of Political Economy 105*(2), 211–48.

Komunjer, I. (2007). Asymmetric power distribution: Theory and applications to risk measurement. *Journal of Applied Econometrics 22*(5), 891–921.

Komunjer, I. and M. Owyang (2011). Multivariate forecast evaluation and rationality testing. *The Review of Economics and Statistics Accepted for publication*(0).

Koop, G., M. Pesaran, and S. M. Potter (1996). Impulse response analysis in nonlinear multivariate models. *Journal of Econometrics 74*(1), 119 – 147.

Korobilis, D. (2009, January). Assessing the transmission of monetary policy shocks using dynamic factor models. Working Paper Series 35-09, The Rimini Centre for Economic Analysis.

Krishnamurthy, A. and Z. He (2011). A model of capital and crises. *Review of Economic Studies 79(2)*, 735–777.

Lai, K. (1990). An evaluation of survey exchange rate forecasts. *Economics Letters 32*(1), 61–65.

Lawrence, M., P. Goodwin, M. O'Connor, and D. Önkal (2006). Judgmental forecasting: A review of progress over the last 25years. *International Journal of Forecasting 22*(3), 493–518.

Litterman, R. (1986). Forecasting with Bayesian vector autoregressions – Five years of experience : Robert B. Litterman, Journal of Business and Economic Statistics 4 (1986) 25-38. *International Journal of Forecasting 2*(4), 497–498.

Liu, P., H. Mumtaz, and A. Theophilopoulou (2011, May). International transmission of shocks: a time-varying factor-augmented VAR approach to the open economy. Bank of England working papers 425, Bank of England.

Lombardi, M., C. Osbat, and B. Schnatz (2012). Global commodity cycles and linkages: a FAVAR approach. *Empirical Economics 43*(2), 651–670.

Ludvigson, S. C. and S. Ng (2009a). A factor analysis of bond risk premia. NBER Working Papers 15188, National Bureau of Economic Research.

Ludvigson, S. C. and S. Ng (2009b). A factor analysis of bond risk premia. NBER Working Papers 15188, National Bureau of Economic Research, Inc.

Ludvigson, S. C. and S. Ng (2009c). Macro factors in bond risk premia. *The Review of Financial Studies 22*(12), 5027–5067.

Marcellino, M., J. H. Stock, and M. W. Watson (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review 47*(1), 1–18.

McCullough, B. (2000). Optimal prediction with a general loss function. *Journal of Combinatorics, Information and System Sciences 25*(1–4), 207–221.

Mishkin, F. (2010). Monetary policy flexibility, risk management, and financial disruptions. *Journal of Asian Economics 21*(3), 242–246.

Nakajima, J. and M. West (2013a). Bayesian analysis of latent threshold dynamic models. *Journal of Business and Economic Statistics 31*, 151–164.

Nakajima, J. and M. West (2013b). Bayesian dynamic factor models: Latent threshold approach. *Journal of Financial Econometrics 11*, 116–153.

Nalewaik, J. (2015, September). Regime-switching models for estimating inflation uncertainty. Technical Report 093, FEDS Working Paper.

Negro, M. D. and C. Otrok (2008). Dynamic factor models with time-varying parameters: measuring changes in international business cycles. Staff Reports 326, Federal Reserve Bank of New York.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica 59*(2), 347–370.

Patton, A. and A. Timmermann (2007a). Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics 140*(2), 884–918.

Patton, A. and A. Timmermann (2007b). Testing forecast optimality under unknown loss. *Journal of the American Statistical Association 102*(480), 1172–1184.

Pesaran, M. and M. Weale (2006). Survey expectations. *Handbook of Economic Forecasting 1*, 715–776.

Philippon, T. (2009). The bond market's q. *The Quarterly Journal of Economics 124*(3), 1011–1056.

Pierdzioch, C., J. Rülke, and G. Stadtmann (2011). On the loss function of the Bank of Canada: A note. *Economics Letters 115*, 155–159.

Pierdzioch, C., J.-C. Rülke, and G. Stadtmann (2013). Oil price forecasting under asymmetric loss. *Applied Economics 45*(17), 2371–2379.

Piger, J., J. Morley, and C.-J. Kim (2005). Nonlinearity and the permanent effects of recessions. *Journal of Applied Econometrics 20*(2), 291–309.

Primiceri, G. E. (2005). Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies 72*(3), 821–852.

Pritsker, M. (2015). Choosing stress scenarios for systemic risk through dimension reduction.

Romer, C. D. and D. H. Romer (1989, March). Does Monetary Policy Matter? A New Test in the Spirit of Friedman and Schwartz. In *NBER Macroeconomics Annual 1989, Volume 4*, NBER Chapters, pp. 121–184. National Bureau of Economic Research, Inc.

Romer, C. D. and D. H. Romer (2004). A new measure of monetary shocks: Derivation and implications. *American Economic Review 94*(4), 1055–1084.

Ruge-Murcia, F. (2003). Inflation targeting under asymmetric preferences. *Journal of Money, Credit and Banking 35*, 763–785.

Schuermann, T. (2014). Stress testing banks. *International Journal of Forecasting 30*(3), 717–728.

Schumacher, C. (2007). Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting 26*(4), 271–302.

Sensier, M. and D. van Dijk (2004). Testing for volatility changes in U.S. macroeconomic time series. *The Review of Economics and Statistics 86*(3), 833–839.

Sims, C. A. (1992a). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review 36*(5), 975 – 1000.

Sims, C. A. (1992b, June). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review 36*(5), 975–1000.

Sims, C. A., D. F. Waggoner, and T. Zha (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics 146*(2), 255–274.

Sims, C. A. and T. Zha (2006, March). Were There Regime Switches in U.S. Monetary Policy? *American Economic Review 96*(1), 54–81.

Stock, J. H. and M. W. Watson (1998). Diffusion indexes. NBER Working Papers 6702, National Bureau of Economic Research.

Stock, J. H. and M. W. Watson (2002a). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association 97*(460), 1167–1179.

Stock, J. H. and M. W. Watson (2002b). Has the business cycle changed and why? *NBER Macroeconomics Annual 17*(1), 159–218.

Stock, J. H. and M. W. Watson (2002c). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics 20*(2), 147–62.

Stock, J. H. and M. W. Watson (2005). Implications of dynamic factor models for VAR analysis. NBER Working Papers 11467, National Bureau of Economic Research.

Svensson, L. (1997). Inflation forecast targeting: Implementing and monitoring inflation targets. *European Economic Review 41*(6), 1111–1146.

Tibshirani, R. (1994). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society, Series B 58*(2), 267–288.

Tran, N. M., M. Osipenko, and W. K. Härdle (2014). Principal component analysis in an asymmetric norm. *arXiv preprint arXiv:1401.3229*.

Tsuchiya, Y. (2016). Assessing macroeconomic forecasts for japan under an asymmetric loss function. *International Journal of Forecasting 32*(2), 233 – 242.

Waggoner, D. F. and T. Zha (1999, November). Conditional Forecasts In Dynamic Multivariate Models. *The Review of Economics and Statistics 81*(4), 639–651.

Weiss, A. A. (1996). Estimating time series models using the relevant cost function. *Journal of Applied Econometrics 11*(5), 539–560.

Weiss, A. A. and A. P. Andersen (1984). Estimating time series models using the relevant forecast evaluation criterion. *Journal of the Royal Statistical Society. Series A (General) 147*(3), 484–487.

West, K., H. Edison, and D. Cho (1993). A utility-based comparison of some models of exchange rate volatility. *Journal of International Economics 35*(1-2), 23–45.

White, H., T.-H. Kim, and S. Manganelli (2015). VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *Journal of Econometrics 187*(1), 169 – 188.

Zarnowitz, V. (185). Rational expectations and macroeconomic forecasts. *Journal of Business and Economic Statistics 3*(4), 293–311.

Zellner, A. (1986). Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association 81*, 446–451.